

Identity and Social Distance in Friendship Formation*

Joan de Martí[†]

Yves Zenou[‡]

March 31, 2011

Abstract

We analyze a network formation model where agents belong to different communities. Both individual benefits and costs depend on direct as well as indirect connections. Benefits of an indirect connection decrease with distance in the network while the cost of a link depends on the type of agents involved. Two individuals from the same community always face a low linking cost and the cost of forming a relationship for two individuals of different communities diminishes with the rate of exposure of each of them to the other community. We derive a number of results with regard to equilibrium networks. In particular, socialization among the same type of agents can be weak even if the within-type link cost is very low and oppositional identity patterns can arise for a wide range of parameters. Our model also suggests that policies aiming at reducing segregation are socially desirable only if they reduce the within-community cost differential by a sufficiently large amount.

Keywords: networks, identity, homophily, social norms.

JEL Classification: D85, J15.

*We are grateful to the participants of the 13th Coalition Theory Network Workshop, the seminars at IFN, IIES, and Duke University, and, in particular, to Andrea Galeotti, Guillaume Haeringer, Rachel Kranton, and Karl Schlag for helpful comments, and to Willemien Kets for a thorough discussion. Joan de Martí gratefully acknowledges financial support from the Spanish Ministry of Education through the project SEJ2006-09993/ECON and a Juan de la Cierva fellowship, co-financed by the European Social Fund, and the support of BarcelonaGSE and the Government of Catalonia.

[†]Universitat Pompeu Fabra and Barcelona GSE. E-mail: joan.demarti@upf.edu.

[‡]Stockholm University, Research Institute of Industrial Economics (IFN), Sweden, and CEPR. E-mail: yves.zenou@ne.su.se.

1 Introduction

The concept of identity has been analyzed for decades in philosophy, psychology, and sociology (see, e.g. Abrams and Hogg, 1999). It is, however, only recently that it has captured the attention of economists. Akerlof and Kranton (2000) were the first to introduce identity into the neoclassical utility maximizing framework in an analysis that draws directly from social psychology.¹

In the present paper, we adopt a different but related view of *identity* by highlighting the importance of the *exposure to the other group* in the friendship formation process between individuals of different ethnic groups. The main novelty in our network formation model is that linking decisions determine the endogenous costs and benefits of individual exposure and identification to other communities.

Motivation. Part of the literature has visualized the concept of identity as *unidimensional*. In other words, individuals with a stronger identification to their own group are usually assumed to have a weaker identification to the other group. Identifications with own and other cultures are treated as *mutually exclusive*. This has usually been studied in societies where a majority and a minority culture coexist. Those who adopt this view consider that ethnic minorities either remain persistent and loyal to their inherited ethnicity or assimilate to the ethnic environment of the majority group. This can lead to the phenomenon of *oppositional identities*, where some ethnic minorities reject the majority behavioral norms while others totally assimilate to it (Ainsworth-Darnell and Downey, 1998). For example, studies in the US (and also in the UK) have found that African American students in poor areas may be ambivalent about learning standard English and performing well at school because this may be regarded as “acting white” (Fordham and Ogbu, 1986; Wilson, 1987; Delpit, 1995; Ogbu, 2003; Austen-Smith and Fryer, 2005; Selod and Zenou, 2006; Battu et al., 2007; Fryer and Torelli, 2010; Battu and Zenou, 2010).²

There is a literature in psychology (see, in particular, Phinney, 1990; Berry, 1997; Ryder et al., 2000) that proposes a broader concept of self-identification in a *two-dimensional* framework, where identifications with two different cultures are not necessary mutually exclusive. Berry (1997) presents four distinct strategies for how individuals relate to two cultures. *Assimilation* is a weak

¹For an overview of the literature on the economics of identity, see Kirman and Teschl (2004) and Akerlof and Kranton (2010).

²There are few theoretical models that try to explain oppositional identity behaviors. Austen-Smith and Fryer (2005) model the trade off experienced by black individuals by putting forward the tension they face between signalling their type to the outside labor market and signalling their type to their peers. Indeed signals that induce high wages can be signals that induce peer rejection. Battu et al. (2007) highlight another trade off faced by blacks. On the one hand, they want to interact with other blacks and thus to reject the white’s norm. On the other, they also want to be friends with whites because the latter possess a better quality social network. They find that black workers can end up choosing oppositional identities if their identity is not strong enough or the wage premium of being employed is high enough. Based on cultural transmission and peer effects, Bisin et al. (2010) develop a dynamic model of identity formation that explains why ethnic minorities may choose to adopt oppositional identities and why this behavior may persist over time.

identification with the culture of origin and a strong identification with the alternative culture. *Integration* is achieved when an individual combines strong dedication to the origin and large commitment to the other culture. *Marginalization* is a weak dedication to both cultures. Finally, *separation* is an exclusive commitment to the culture of origin. The following figure summarizes these four different possibilities in a two-dimensional space.

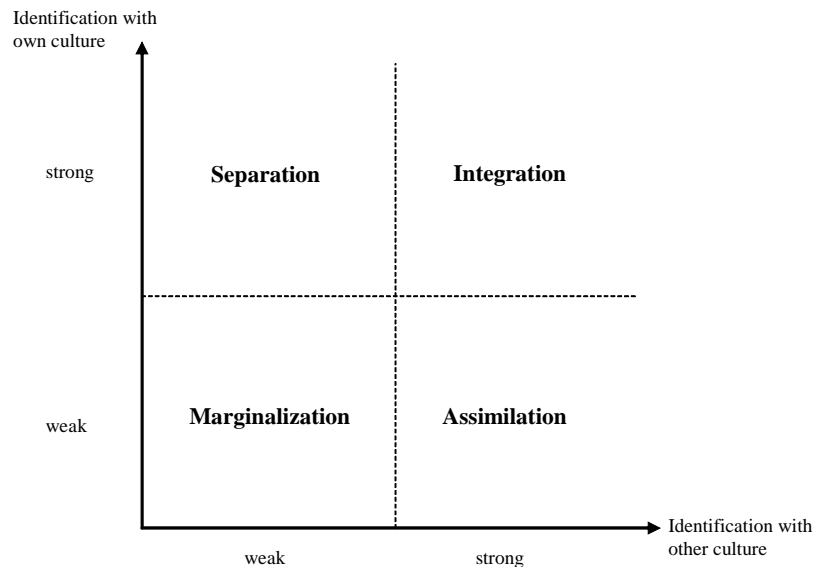
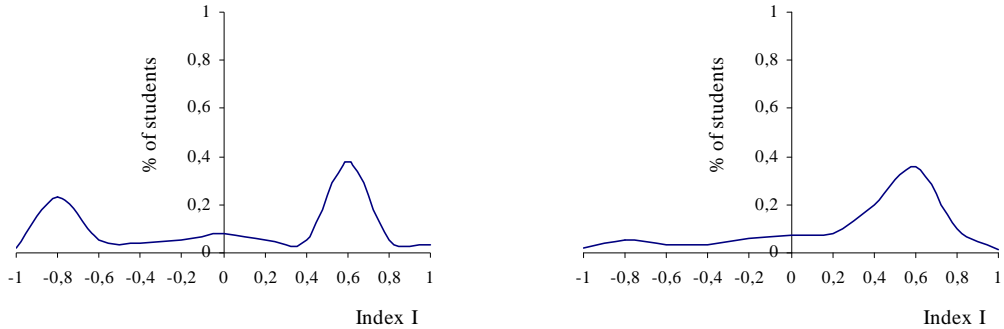


Figure 1. Different identifications for ethnic minorities

As it can be seen from Figure 1, individuals who are integrated have not only a strong identification to the majority culture but also to their own culture. Observe that the concept of oppositional identity corresponds to either a separated or an assimilated individual in Figure 1.

There are some empirical studies in the US using both the unidimensional and bidimensional definition of identity choices. For example, using the National Longitudinal Study of Adolescent Health (AddHealth), Patacchini and Zenou (2006) use the homophily index H_i of individual i proposed by Coleman (1958) to analyze the exposure of individuals of white and black race to own and other races. If the homophily index H_i of a student i is equal to 0, it means that the percentage of same-race friends of this individual equals the share of same-race students in the school. Negative values of the index imply an underexposure to same race students, while positive values imply an overexposure to same race students compared to the mean. Figure 2 displays their results for mixed schools (i.e. schools with a percentage of black and white students between 35 and 75 percent).



Blacks in integrated schools Whites in integrated schools
 Figure 2. Distribution of students by share of same-race friends in integrated schools

Most of white students have white friends since roughly 40 percent of them are associated with values of the homophily index greater than 0.4, denoting a clear deviation from the assumption of random choice of friends by race. Black students appear to be more heterogeneous in their choice of friends than whites. The clear bimodality in the distribution (corresponding to values of H_i between -0.6 and -0.8 and between 0.6 and 0.8) reveals that there are, mainly, two types of black students: those who have mostly white friends and those choosing mostly black friends. In terms of Berry’s characterization presented above (Figure 1), most white students and some black students show a separated or integrated identities, while a relevant fraction of black students shows assimilated identities.³

A model of homogeneous behavior among members of the same groups cannot explain the pattern obtained in Figure 2. Choices of friends between races need to be consistent with each other in order for the observed aggregated level of social interactions to show the emergence of heterogeneous identity patterns. Thus, to understand the observed patterns, the *network* aspect of friendships cannot be ignored.

Model and Results. We propose a network formation model that can simultaneously explain the identity patterns described in Figure 1 and the socialization patterns observed in Figure 2.. We consider a finite population of individuals composed by two different communities. These two communities are categorized according to some exogenous factor such as, for example, their gender, race or ethnic and cultural traits. Individuals decide with whom they want to form a link with according to a utility function that weights the costs and benefits of each connection. This results in a network of relationships where a link between two different individuals represents a friendship relationship. The utility of each individual depends on the geometry of this friendship network.

To model the benefits and costs of a given network, we consider a variation of the connections

³Marmaros and Sacerdote (2006) show that the main determinants of friendship formation are the geographical proximity and race. Also Mayer and Puller (2008), using administrative data and information from Facebook.com, find that race is strongly related to social ties, even after controlling for a variety of measures of socioeconomic background, ability, and college activities.

model introduced by Jackson and Wolinsky (1996), a workhorse model in the analysis of strategic network formation.⁴ From the standard connections model, we keep the property that an individual benefits from her direct and indirect connections, and that this benefit decays with distance in the network. This can be interpreted as positive externalities derived from information transmission (of trends and fashion for adolescents, of job offers for workers, etc.). However, in the standard connections model, each link is equally costly, irrespective of the pair of agents that is connected. We depart from this assumption as follows.

Consider the case where communities are defined according to ethnicity, which may entail differences in language and social norms. When two individuals of different communities interact, they may initially experience a disutility due to the attachment to their original culture. This discomfort can, however, be mitigated if individuals are frequently exposed to the other community. Indeed, when someone spends time interacting with people from the other community, she can learn the codes and norms (prescriptions) that govern their social interactions. This is precisely the starting point of our analysis: the *exposure* to another social group decreases the cost of interacting with individuals from that group.

To be more precise, we assume that the linking cost of a pair of agents belonging to different communities depends on their level of exposure to the other community. We model this feature through a cost function that *positively* depends on the fraction of same-type friends each person has. This cost is, however, never lower than the intracommunity linking cost.

In this respect, social distance expresses the force underlying this cost structure. Two agents are closer in the social space, the more each of them is exposed to the other community. And, the closer they are in the social space, the easier it is for them to interact. In our model, *this social distance is endogenous* and depends on the respective choice of peers.

We study the shape of stable networks in this setup. We use the notion of pairwise stability, introduced by Jackson and Wolinsky (1996). It is a widespread tool in the strategic analysis of social and economic networks. It takes into account the *individual incentives* to create and sever links and the necessary *mutual consent* between both sides for a link to be formed. In a nutshell, a network is pairwise stable if no agent has incentives to sever any of her links, and no pair of agents who are not connected have incentives to form a new link. In our model, it is a complex combinatorial problem to fully characterize the set of stable networks. We provide, however, a partial characterization that conveys information about the different identity and socialization patterns that may arise in equilibrium.

In this context, when *intracommunity* linking costs are low, we show that oppositional identities can emerge when *intercommunity* costs are also low, i.e. the maximum possible cost of an intercommunity link is close to the cost of an intracommunity link. In several equilibrium configurations, *bridge links* (i.e. links that connect both communities) prevail. Even if those bridge links can be

⁴See Goyal (2007) and Jackson (2008) for overviews of the growing literature on social and economic networks.

quite costly for the agents involved, these links give them direct access to parts of the networks that would be not accessible otherwise. This reverberates into direct and indirect benefits that overcome the cost for both sides of the link, and acts as positive externalities for the agents who are in their respective neighborhoods since the cost of a link is only paid by the individuals directly involved in it. We can also determine conditions under which totally *assimilated* and *separated* minorities (Figures 1 and 2) can emerge in equilibrium as well as “extreme” networks (i.e. bipartite networks) where individuals of each community are only connected to individuals from the other community.

The mechanism we suggest links *socialization costs* with *network geometry*. Since individual and aggregate welfares depend on the geometry of the resulting network, we may wonder about the impact of policies aiming at reducing intercommunity socialization costs. In our context, such an analysis is difficult to perform due to the inherent multiplicity of stable configurations. We try, however, to perform one step in this direction by comparing two extreme outcomes: extremely integrated and segregated networks. When intracommunity costs are low, we show that social integration is not always preferred to social segregation. The inefficiency comes from the excessive individual cost paid to build bridge links between communities. This suggests that these types of policies may only be effective if they substantially reduce intercommunity socialization costs. We believe that this is an interesting result that may explain part of the relative inefficiency of integration policies such as school busing, forced integration of public housing, and Moving to Opportunity (MTO) programs implemented in the United States (the latter relocate families from high- to low-poverty neighborhoods (and from racially segregated to mixed neighborhoods)).⁵ In our theoretical framework, policies reducing intercommunity socialization costs are not necessarily going to induce more desirable network structures. For example, activities outside the classroom for adolescents or cultural activities at the neighborhood level can favor integrated patterns since they may facilitate interactions among individuals of different types, but the outcome is not going to be socially efficient unless these policies sufficiently decrease the cost of interactions.

Our model can be extended in a number of directions. We present two different possible extensions in the last section of the paper. First, we introduce heterogeneous payoff externalities. It might be that agents of one type exert a higher direct positive externality on others than agents of the other type. This setup can represent, for example, a situation in which one of the two types has *ex ante* a higher human and/or social capital.⁶ Second, we introduce a social punishment for individuals from the minority group who identify themselves with the majority culture. This punishment expresses the rejection by the members of her original group who strictly stick to their social and cultural values. This can be a reduced form representation of the “acting white” phenomenon mentioned above. We show that both situations facilitate the adoption of oppositional identities.

⁵See Lang (2007), which gives a very nice overview of these policies in the United States.

⁶Benabou (1996) studies a location model with two types and heterogeneous human capital externalities with a similar feature.

Related Literature. The papers by Currarini et al. (2009, 2010), Bramoullé and Rogers (2010), and Mele (2010) study homophily in networks using stochastic models of network formation. The aim in these papers is therefore similar, but there are important differences with respect to the methodology. They assume a dynamic and stochastic matching sequence while we study strategic linking decisions in a one-shot game. The papers by Currarini et al. (2009, 2010) develop a matching model with a population formed by communities of different sizes. They are able to replicate a number of observations from real-world data related to homophilous behavior at the aggregate level but, in their model, individuals' behavior is totally homogeneous among the same group of agents. Bramoullé and Rogers (2010) depart from Currarini et al. by assuming that dynamic matching follows the process studied by Jackson and Rogers (2007) and they show that more connected individuals tend to have a more diverse set of friends.⁷ Mele (2010) studies a model where meetings are dynamic and stochastic and each individual involved in a meeting can decide whether she wants to create or sever the link with the other person. Mele shows that this process always converges.

The papers by Johnson and Gilles (2000) and Jackson and Rogers (2005) extend the Jackson and Wolinsky (1996)'s connection model by introducing ex ante heterogeneity in the cost structure. In the latter model, the cost of creating links between the two communities is exogenous and does not depend on the behavior of the two agents involved in the connection. In the former model, the cost of creating a link is proportional to the geographical distance between two individuals and thus this cost is fixed ex-ante and does not change with the linking decisions of the two agents involved in the link.⁸ This turns out to be a key difference with our cost structure, where the cost of a link is endogenous and depends on the neighborhood structure of the two agents involved in the link.

Eguia (2010) presents a theory in which the cost of assimilation is endogenous and strategically chosen by the better-off group in order to screen those who wish to assimilate. Eguia (2010) shows that, in equilibrium, only high types who generate positive externalities to the members of the better-off group will assimilate. The paper does not focus on network issues and therefore the results are of a different and complementary nature.

Some papers analyze the consequences of homophily in social networks. For example, Buhai and Van der Leij (2008) develop a social network model of occupational segregation with inbreeding bias, and Golub and Jackson (2008) study how homophilous networks affect communication and agents' beliefs in a dynamic information transmission process.

Finally, Schelling (1971) is a seminal reference when discussing social networks and segregation patterns. Shelling's model shows that, even a mild preference for interacting with people from the same community can lead to large differences in terms of location decision. Indeed, his results suggest that total segregation persists even if most of the population is tolerant about heteroge-

⁷This holds when students are divided according to sex in high-schools in the AddHealth data set.

⁸Iijima and Kamada (2010) extend the model of Johnson and Gilles (2000) to a multidimensional setting where individuals care for the exogenous distance of different attributes at the same time.

neous neighborhood composition.⁹ Our analysis differs from Schelling’s classical framework (and its different extensions) in several directions. First of all, we analyze a network formation game while in Schelling the network structure is fixed. Secondly, homophilous preferences in our setup are not homogenous and are endogenous. In particular, these preferences are determined by both the direct and indirect benefits derived from the creation of a link and by the social environment of the potential partner. The economic benefits thus depend on the network structure of all the population.

Our main contribution. Our main contribution is to show that the mechanism of our model (that relates the cost of friendship to the social distance of two linked individuals) can induce endogenous asymmetric socialization behaviors of a particular, and economically relevant, type. We assume that socialization costs depend on exposure to other communities and we show that *ex ante identical individuals* may end up with very different network positions. In particular, *separated, integrated, marginalized* and/or *assimilated* patterns of friendships (Figure 1) may prevail in equilibrium. Thus, we obtain intragroup asymmetric behaviors in connectivity in a number of equilibrium networks, which allow us to rationalize the friendship patterns observed in Figure 2. We do not mean here that the result of socialization is always going to lead to segregation and/or oppositional identities, but we are able to show that these patterns can emerge in some circumstances as the result of a *decentralized* process of socialization. There are also other possible equilibria where this would not occur and our direct aim *is not to provide a full characterization of the set of equilibrium networks*. Indeed, the pool of high-schools from the AddHealth data set shows a variety of real-world configurations. Therefore, it is natural that any model that wants to give reasonable microfoundations for these configurations exhibits multiplicity of equilibria. We endogenously model the structure of the network of friendship relations where not only friends, but friends of friends, and friends of friends of friends, etc. matter. Because of this feature, a problem of a combinatorial nature, also present in the classical model of Jackson and Wolinsky (1996), emerges.¹⁰ This is why it is extremely hard, if not impossible, to provide a full-fledged characterization of all possible stable networks.¹¹

⁹This framework has been modified and extended in different directions, exploring, in particular, the stability and robustness of this extreme outcome (see, for example, Mobius, 2007 or Zhang, 2004).

¹⁰It is indeed well-known that non-cooperative games of network formation with nominal lists of intended links are plagued by coordination problems (Myerson, 1991; Jackson, 2008; Cabrales et al., 2011). Cooperative-like stability concepts solve them partially, but heavy combinatorial costs still jeopardize a full characterization.

¹¹The existence of a plethora of equilibria in our framework is not the result of the use of a weak stability concept (in our case, pairwise stability). The use of an stronger equilibrium concept in network formation games, such as Pairwise Nash equilibria, does not seem to significantly reduce the number of equilibria: in a slightly perturbed version of the present model, we are able to show that the set of pairwise stable equilibria and the set of pairwise Nash equilibria coincide. This is available upon request.

2 The model

2.1 Individuals, communities, and networks

There is a finite population of individuals denoted by $N = \{1, \dots, n\}$. This population is divided into two communities, the *Blue* and the *Green* communities. Each agent belongs exclusively to one of the two communities, B or G . This initial endowment of each individual can be interpreted, for example, as the identity inherited from her family. The type of individual i is denoted by $\tau(i) \in \{B, G\}$. Let n^B denote the number of B individuals in the population. Similarly, let n^G denote the number of G individuals in the population. We have that $n = n^B + n^G$. We assume, without loss of generality, that $n^B \leq n^G$.

Individuals will be connected through a social network structure. A network is represented by a graph, where each node represents an individual and a connection among nodes represents a friendship relationship between the two individuals involved. We denote a network by g , and $g_{ij} = 1$ if i is friend with j and $g_{ij} = 0$ otherwise. In our framework, friendship relationships are taken to be reciprocal, i.e. $g_{ij} = g_{ji}$ so that graphs/networks are *undirected*. We denote the link of two connected individuals, i and j , by ij . The set of i 's direct contacts is: $N_i(g) = \{j \neq i \mid g_{ij} = 1\}$, which is of size $n_i(g)$. The direct contacts of individual i of the same type is $N_i^{\tau(i)}(g) = \{j \neq i, \tau(i) = \tau(j) \mid g_{ij} = 1\}$, and we denote the cardinality of this set by $n_i^{\tau(i)}(g)$.

We present some examples of network configurations. The *circle* is such that each agent has two direct contacts. The *star-shaped network* has one central agent who is in direct contact with all the other peripheral agents who, in turn, are only connected to this central agent. The *complete network* is such that each agent is in direct relationship with every other agents so that each individual i has $n - 1$ direct contacts.

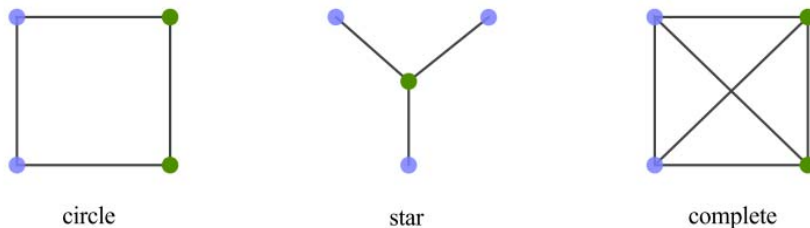


Figure 3. Circle, star and complete networks with four individuals.

A network is depicted as a set of colored nodes (Figure 3), which allows to distinguish among members of different groups, and links that connect some or all of them. Naturally, blue nodes refer to type- B individuals while green nodes indicate type- G individuals.

The circle and the complete network are examples of regular configurations in which all agents share a similar position, though they differ by the number of connections each agent possesses. The star is an example of centralized, asymmetric, network structure, where the center occupies a very different position than the rest of the other individuals in the network.

We still need to introduce some more concepts associated to the connectivity of the network.

There is a *path* in network g from individual i to individual j if there exists an ordered set of individuals, with i being the first one and j being the last one, such that each agent is connected to the following one according to this order.¹² Graphically, there is a path from individual i to individual j whenever one can travel from i to j through the links of the network. The length of a path is the number of links involved in it. The *shortest path* between from i to j is the path that involves the lowest number of links. We define the *geodesic distance* (or simply distance) between individuals i and j as the length of the shortest path that connects them, and we denote it by $d(i, j)$. If in a given network there does not exist any path that connects individuals i and j we say that the distance between them is infinite, and $d(i, j) = \infty$. For example, in a star-shaped network any two different agents in the periphery are connected by a path of distance two. Since there is no other shorter path that connects these two peripheral agents, the distance among them in the network is equal to two. Finally, we say that a link among individuals i and j is a *bridge link* whenever these two individuals are of different types. Formally, the link ij is a bridge link if $\tau(i) \neq \tau(j)$. Bridge links are the ones that connect both communities.

2.2 Preferences

The utility function of each individual i , denoted by $u_i(g)$, depends on the network structure that connects all the population. It is given by

$$u_i(g) = \sum_j \delta^{d(i,j)} - \sum_{j \in N_i(g)} c_{ij}(g) \quad (1)$$

where $0 \leq \delta < 1$ is the benefit from links, $d(i, j)$, the *geodesic distance* between individuals i and j , and $c_{ij} > 0$ is the cost for individual i of maintaining a direct link with j .

The utility function (1) has the general structure of the so-called *connections model*, introduced by Jackson and Wolinsky (1996). Links represent friendship relationships between individuals and involve some costs. A “friend of a friend” also results in some indirect benefits, although of a lesser value than the direct benefits that come from a “friend”. The same is true of “friends of a friend of a friend,” and so forth. The benefit deteriorates in the geodesic distance of the relationship. This is represented by a factor δ that lies between 0 and 1, which indicates the benefit from a direct relationship between i and j , and is raised to higher powers for more distant relationships. For

¹²Formally, a *path* p_{ij}^k of length k from i to j in the network g is a sequence $\langle i_0, i_1, \dots, i_k \rangle$ of players such that $i_0 = i$, $i_k = j$, $i_p \neq i_{p+1}$, and $g_{i_p i_{p+1}} = 1$, for all $0 \leq p \leq k - 1$, that is, players i_p and i_{p+1} are directly linked in g . If such a path exists, then individuals i and j are path-connected.

instance, in the network described in Figure 4, individual 1 obtains a benefit of 2δ from the direct connections with individuals 2 and 3, an indirect benefit of δ^2 from the indirect connection with individual 4, and an indirect benefit of $2\delta^3$ from the indirect connection with individuals 5 and 6. Since $\delta < 1$, this leads to a lower benefit of an indirect connection than of a direct one.

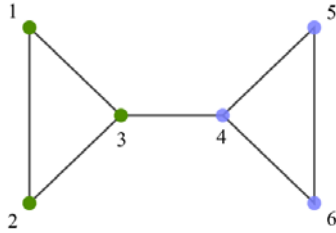


Figure 4. A bridge network.

However, individuals only pay costs $c_{ij} > 0$ for maintaining their *direct* relationships. This is where our model becomes very different from the standard connections model. To characterize linking costs, we need first to introduce one more concept. Given a network g , we define the *rate of exposure* of individual i to her own community $\tau(i)$ as:

$$e_i^{\tau(i)}(g) = \frac{n_i^{\tau(i)}(g)}{n_i(g) - 1}. \quad (2)$$

This ratio $e_i^{\tau(i)}(g)$ measures the fraction of *same-type friends* since $n_i^{\tau(i)}(g)$ is the number of i 's same-type friends in network g while $n_i(g)$ is the total number of i 's friends in network g , independently of their type. The reason why we subtract a 1 in the denominator will become apparent in the next paragraphs.

We can now introduce the cost structure. Let c and C be strictly positive constants. We assume that:

$$c_{ij}(g) = \begin{cases} c & \text{if } \tau(i) = \tau(j) \\ c + e_i^{\tau(i)}(g)e_j^{\tau(j)}(g)C & \text{if } \tau(i) \neq \tau(j) \end{cases} \quad (3)$$

There are thus different costs, depending with whom a connection is made. Since $C > 0$ and rate of exposures are non-negative, the main feature of this cost structure is that it is always more costly to form a friendship relationship with someone from the other community (the cost is given by (3)) than with someone from the same community (which cost is c). In particular, if an individual i of type $\tau(i)$ forms a friendship relationship with an individual j of type $\tau(j)$, with $\tau(i) \neq \tau(j)$ (i.e. intercommunity friendship formation), then the cost is increasing in their respective rates of exposure to their own communities. If, for example, a green person has only green friends, then it will be difficult for her to interact with a blue person, especially if the latter has mostly blue

friends. There are different cultures, norms and habits between communities so that frictions are higher the more different people are. If we interpret “type” by “race” so that “blue” and “green” are replaced by “black” and “white”, then (3) means that it is always easier for blacks to interact with other blacks and likewise for whites, and that the interracial relationships strongly depend on how “exposed” individuals are, i.e. how many same-race friends they have. These difficulties in interracial relationships can be due to languages issues¹³ or more generally to different social norms and cultures.^{14,15}

What we have in mind here is that individuals are born with a certain type τ (blue or green) that affects their easiness to interact with other individuals. It is assumed that it is less costly to interact with someone of the same type than of a different type. So from this initial trait τ , there are natural gaps and differences between communities of types. But people make choices in terms of friendships, and that can be interpreted in terms of identity. These choices can increase or decrease the original gap between individuals. If someone who is born blue chooses to have only blue friends (this is an identity choice), then it will be more difficult for her to interact with a green person. However, the more similar the choices are, the easier it is to interact with someone from a different type. Observe that we allow that friend choices can totally erase the initial cost gap between a blue type and a green type. Indeed, if at least one individual (i or j) has no friends of the same type (i.e. $e_i^{\tau(i)} = 0$ or $e_j^{\tau(j)} = 0$), then it is equally costly for them to interact with each other than with someone of same type (i.e. the cost is c in both cases).¹⁶

The reason why we subtract a 1 in the denominator in the definition of the rate of exposure (see (2)) is because, when we compute the cost of a given bridge link between communities, we do not include this bridge link in the computation of the cost. What is relevant for the cost is the rate of exposure according to the rest of connections of each of the two individuals involved in the

¹³For example, the studies of Labov (1972), Baugh (1983), and Labov and Harris (1986) reveal that Black English of different metropolitan areas has converged, while it has been simultaneously diverging from Standard American English. This creates some costs in the interactions between blacks and whites.

¹⁴Camargo et al. (2010) show in a randomized experiment that whites who are randomly assigned black roommates have in the future a significantly larger proportion of black friends than white students who are randomly assigned white roommates. Ben-Ner et al. (2009) show in lab experiments that the distinction between in-group and out-group affects significantly economic and social behavior, for example, in forming working relationships.

¹⁵Lemanski (2007) documents an interesting experiment in post-apartheid urban South Africa by examining the lives of those already living in desegregated spaces. She studies the case a low-cost state-assisted housing project situated in the wealthy southern suburbs of Cape Town. In this social housing project, named Westlake village, colored and Black African (alongside a handful of white and Indian) residents were awarded state housing in 1999 as replacement for their previous homes, which were demolished to make way for a mixed land-use development. She find that different races are not only living peacefully in shared physical space but also actively mixing in social, economic and to a lesser extent political and cultural spaces. Furthermore, residents have largely overcome apartheid histories and geographies to develop new localized identities. This can be another indication that when people from different races or cultures interact with each other the costs of further interacting decreases.

¹⁶In Appendix B, we investigate a different cost function where the intercommunity cost is *not* anymore equal to the intracommunity even if one of the persons involved in a relationship has no friends of the same type.

bridge link.

To illustrate our cost function (3), consider again the network described in Figure 4 and assume that individuals 1, 2, and 3 are greens (type G) while individuals 4, 5, and 6 are blues (type B). Imagine that individuals 3 and 4 are not yet connected and individual 3 considers the possibility of creating a link with 4. In that case, the cost of connecting 3 (green) to 4 (blue) is:

$$c_{34}(g) = c + \frac{n_3^{\tau(3)}(g)}{n_3(g) - 1} \frac{n_4^{\tau(4)}(g)}{n_4(g) - 1} C = c + C$$

since $n_3^{\tau(3)}(g) = n_4^{\tau(4)}(g) = 2$ (number of same-type friends of 3 and 4, respectively) and $n_3(g) = n_4(g) = 3$ (total number of 3's and 4's friends independently of type, considering also the link between them),¹⁷ which implies that $e_3^{\tau(3)}(g) = e_4^{\tau(4)}(g) = 1$.

If, for example, individual 4 also had a link with 2, the cost of connecting 3 (green) to 4 (blue) would be

$$c_{34}(g) = c + \frac{n_3^{\tau(3)}(g)}{n_3(g) - 1} \frac{n_4^{\tau(4)}(g)}{n_4(g) - 1} C = c + \frac{2}{3}C$$

since $e_3^{\tau(3)}(g) = 1$ but $e_4^{\tau(4)}(g) = 2/3$. It would be less costly for individual 3 (green) to be friend to individual 4 (blue) in this situation because the latter has already a green friend.

With the above notation, we want to highlight that costs, in particular intercommunity costs, depend on the network structure. However, from now on, and to minimize notational burden, we will not make the dependency of the rates of exposure and the linking costs on g explicit.

2.3 Network stability

In games played on a network, individuals payoffs depend on the network structure. In our case, this dependency is established in expression (1), that encompasses both the benefits and costs attributed to an individual given her position in the network of relationships. Any equilibrium notion introduces some stability requirements. The notion of pairwise-stability, introduced by Jackson and Wolinsky (1996), provides a widely used solution concept in networked environments. Let us now define this concept.

Definition 1 *A network g is **pairwise stable** if and only if:*

(i) *for all $ij \in g$, $u_i(g) \geq u_i(g - ij)$ and $u_j(g) \geq u_j(g - ij)$*

(ii) *for all $ij \notin g$, if $u_i(g) < u_i(g + ij)$ then $u_j(g) > u_j(g + ij)$.*

¹⁷Observe that, when individual 3 considers the possibility of creating a link with individual 4, individual 3 does not take into account the possible link between 3 and 4 when calculating the percentage of same-race friends of herself and of 4.

In words, a network is pairwise-stable if (i) no player gains by cutting an existing link, and (ii) no two players not yet connected both gain by creating a direct link with each other. Pairwise-stability thus only checks for one-link deviations.¹⁸ It requires that any mutually beneficial link be formed at equilibrium but does not allow for multi-link severance.

We will use throughout this equilibrium concept. Thus, network g is an equilibrium network whenever it is pairwise stable.

3 Stable networks

3.1 Low intra-community costs

We start the analysis of stable networks with the case of low intra-community costs c . In particular, we start assuming that $c < \delta - \delta^2$. If there were only one community (i.e. only one type of individuals), then the complete network would be the unique equilibrium network (as in the connections model of Jackson and Wolinsky, 1996). But, since we have two different communities and different cost structures, this is not anymore true. Indeed, an individual of one type may decide to lower the exposure to her own community in order to become more attractive to the other one. We start this section by trying to understand under which conditions this may not happen so that we can still obtain fully intraconnected communities.

We use the following definitions: A network displays *complete integration* when both communities are completely connected, *complete segregation* when both communities are isolated and *partial integration* in any other case. We have the following result:¹⁹

Proposition 1 *Assume*

$$c < \delta - \delta^2 \tag{4}$$

so that that each community is fully intraconnected. Then,

- (i) *The network such that the blue and the green communities are **completely integrated** is an equilibrium network if and only if*

$$C \leq \frac{(n-2)^2 (n-3)}{n^G (n^G - 1)^2} (\delta - \delta^2 - c). \tag{5}$$

- (ii) *If*

$$C > \delta + (n^B - 1) \delta^2 - c \tag{6}$$

*holds, then the network for which the blue and the green communities are **completely segregated** is an equilibrium network.*

¹⁸This weak equilibrium concept is often interpreted as a necessary conditions for stronger stability concepts.

¹⁹All proofs can be found in the Appendix.

To interpret these results, it is useful to think about two different effects. The first effect, which we refer to as the *connections effect*, expresses the role of the direct and indirect gains and losses of forming or severing a link. This first effect is also present in the connections model of Jackson and Wolinsky (1996). There is, however, a second effect, that we refer to as the *exposure effect*, which is new. This effect is due to the fact that the formation a new link affects the exposure rates of the individuals involved in it. Indeed, if the new link is between two individuals from different communities (the same community), then the rate of exposure of each of these individuals to their own community is going to decrease (increase) and thus their intercommunity costs will decrease (increase). This indirect exposure effect is positive (negative).

As a result, the completely integrated network is going to be stable if the sum of the connections and the exposure effects for any link is positive. Consider an intercommunity link. The connections effect is ambiguous because the cost of keeping the link for each individual is strictly large than c since their rates of exposure to their own communities are strictly positive. However, severing such a link has an strong and negative *exposure effect* since it increases both their rate of exposure and the intercommunity costs with the rest of their friends. Some algebra shows that this second (exposure) effect always dominates the connections effect and, hence, nobody has incentives to sever a link. The case of an intracommunity link is less clear. In such a case, the connections effect is clearly signed: it is positive because we are assuming that $\delta - \delta^2 - c > 0$, which implies that for two individuals from the same community the benefits of a direct connection compared to an indirect connection of distance two always outweigh the costs of forming such a link. However, keeping such link has a negative exposure effect: it increases their respective rates of exposure to their own communities, and therefore the costs of their intercommunity links become larger. If C is sufficiently large, then the negative exposure effect dominates the positive connections effect and we end up with a stable completely integrated network (see (5)).

The completely segregated network arises when the connections effect of an intercommunity link is negative. Condition (6) is precisely the mathematical formulation of this negative effect. Note that, in this case, there are no exposure effects to consider since we start from a situation where there are no intercommunity links.

If we now use the two-dimensional definition of identity, illustrated in Figure 1, the blues and greens are here *separated*. This could be a case where the two populations are physically separated (i.e. spatially segregated) so that interactions are very costly (because, for example, of commuting costs, prejudices, etc.). Intuitively, if C decreases, individuals may start forming bridge links. These links may make them more attractive, because of the exposure effects, to the other community members, who, in turn, form bridge links, etc. Let us investigate in more details this *partially-integrated* case, where there are some bridges between both communities.

Define

$$\Phi(n^\tau, \delta, c) \equiv \frac{n^\tau [\delta + (n^\tau - 2)\delta^2 - (n^\tau - 1)\delta^3 - c]}{n^\tau - 1}$$

The following proposition characterizes some partially integrated equilibrium networks, and bring into the picture a third important component in the stability of a network geometry:

Proposition 2 *Assume (4).*

(i) *If*

$$C > \max \left\{ \frac{n^G (\delta - \delta^2 - c)}{n^G - 2}, \Phi(n^G, \delta, c), \Phi(n^B, \delta, c) \right\} \quad (7)$$

holds, then the network where both communities are fully intraconnected and where there is only one bridge link is an equilibrium network (Figure 5).

(ii) *If*

$$\frac{n^G n^B (\delta - \delta^2 - c)}{(n^G - 1)(n^B - 1) - n^B} < C < \delta - \delta^3 - c \quad (8)$$

holds, then the network where both communities are fully intraconnected and each blue individual has one, and only one, bridge link and where each green individual has at most one bridge link is an equilibrium network (Figure 6).

(iii) *If*

$$\frac{(\delta - \delta^2 - c) n^G}{(n^G - 1)} < C < \frac{n - 2}{n^B - 1} \min \left\{ \frac{n - 3}{n^B - 2} (\delta - \delta^2 - c), (1 - \delta) (\delta + (n^B - 1) \delta^2) - c \right\}$$

holds, then the network in which both communities are fully intraconnected and only one blue agent connects with the green community by linking to all green individuals is an equilibrium (Figure 7).

In these equilibrium configurations some integration between blues and greens is taking place. The following figures provide a graphical representation.

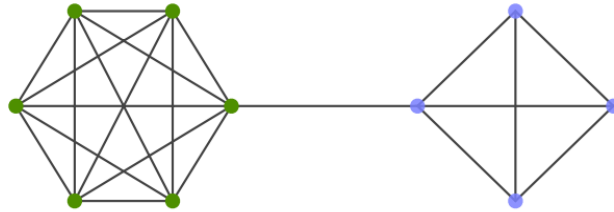


Figure 5. Equilibrium network when condition (7) holds.

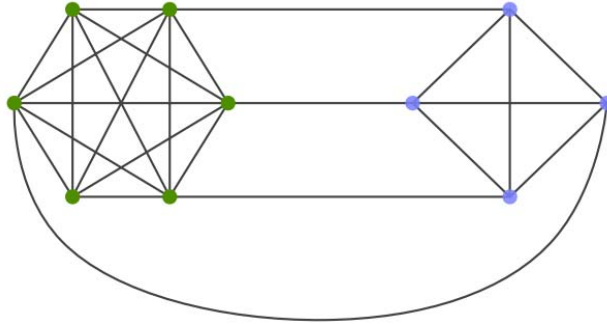


Figure 6. Equilibrium network when condition (8) holds.

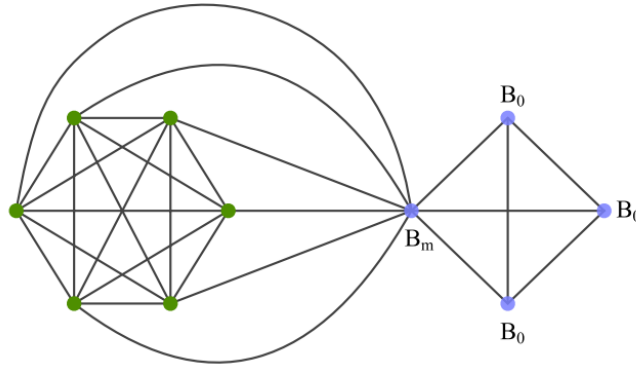


Figure 7. Oppositional identities when $c < \delta - \delta^2$.

As before, both connections and exposure effects are present in explaining the results of Proposition 2. There is, however, a third component that becomes more relevant here: the requirement of *mutual consent* for the link to be formed. When the intercommunity costs of forming a link are relatively large, i.e. C is high, the network in Figure 5 is pairwise stable because the connections effect for the agents involved in the only bridge link between communities is positive while the connections effect of any other intercommunity link is negative for at least one of the two sides of each of these potential links.²⁰

When C decreases slightly, individuals from different communities may now want to create one of these missing links. This is illustrated in Figure 6. While the direct benefits of a such

²⁰The two individuals involved in this bridge link enjoy a singular position in the network. Some literature in sociology has highlighted the importance of these type of links in terms of social capital: it is important that bridges exist between communities. Indeed, social capital is created by a network in which people can broker connections between otherwise disconnected segments (Granovetter, 1973, 1974; Burt, 1992). We can say that the people who are bridging two communities are sitting in a *structural hole* of the network. A structural hole exists when there is only a weak connection between two clusters of densely connected people (Burt, 1992; Goyal and Vega-Redondo, 2007).

new connection have not changed, the costs are now reduced and, as a result, the sign of the connections effect of such a new link is reverse. In both networks described in Figures 5 and 6, the exposure effects play no role since each of the agents is involved in at most one link and the cost of this link is kept constant when there are changes in the connections within the community (these intracommunity links do not change the rate of exposure of individuals, which remains maximal and equals to 1, according to the definition of rate of exposure given in (2)).

The logic behind the stability of the network displayed in Figure 7 is different since it strongly relies on the exposure effect. The B_m blue individual invests in a large number of intercommunity links in order to decrease enough her own rate of exposure, and thus to decrease her own cost of each of these connections. This, in turn, make it cheaper for each green individual to connect to her and win direct access to the blue community.

To understand our results, let us summarize the three main forces at work:

(1) Individuals want to form connections to obtain direct and indirect benefits. In a disperse network, connecting to a member of a different community usually gives access to many opportunities. This is the *connections effect*.

(2) Because links are costly, individuals become more attractive the more they are friends with individuals from the other community and hence can form new links more easily with the other community. This is the *exposure effect*.

(3) There is a *coordination problem* because the creation of a link needs the consent of both individuals. Condition (i) in Definition 1 of pairwise stability highlight this *mutual consent effect*.

Equilibrium networks are those that correctly balance these three forces at the individual level. The equilibrium networks characterized in Proposition 1 and 2 provide some understanding on how these three effects interact with each other. Contrary to the literature on segregation (e.g. Schelling, 1971; Benabou, 1993) and on friendship formation (Austen-Smith and Fryer, 2005; Battu et al., 2007), it is important to observe that both the *individual location* and the *structure of the network* are here crucial to understand the equilibrium outcomes. Indeed, not only benefits but costs are affected by individual's location and the structure of the network. For example, two identical blue individuals who have different positions in the network may have different incentives to form a link with a green person so that, in equilibrium, only one of them will find it beneficial to form a bridge link.

Let us now investigate the issue of *assimilation* and *oppositional identities*. The completely integrated network in Proposition 1 provides a first example of how our model can generate some assimilation patterns. For example, blue individuals partially assimilate to the green community because each blue individual interacts more often with green individuals than with blue individuals. However, in such a case the reason mostly relies on the difference in size of both communities. A more interesting and richer example is the network displayed in Figure 7. Here, the three effects mentioned above play a role to generate the assimilation pattern of individual B_m . Assimilation

arises because the exposure effect of B_m (blue) individual increases the magnitude of the connections effect for each of the green individuals, which, in turn, induces mutual consent.

With regard to individual welfare, it is not always true that oppositional blue individuals obtain a higher utility than non-oppositional blues. Take, for example, Proposition 2 (*iii*) (Figure 7). The equilibrium utility of the oppositional blue B_m is

$$U_{B_m} = (n - 1) (\delta - c) - n^G \left(\frac{n^B - 1}{n - 1} \right) \left(\frac{n^G - 1}{n^G} \right) C$$

while the utility of non-oppositional blues is:

$$U_{B_0} = (n^B - 1) (\delta - c)$$

So we have

$$U_{B_m} \gtrless U_{B_0} \Leftrightarrow C \lesseqgtr \frac{n^G}{n^G - 1} \frac{n - 1}{n^B - 1} (\delta - c)$$

This inequality is not incompatible with the condition given in Proposition 2 (*iii*), meaning that both cases, $U_{B_m} > U_{B_0}$ and $U_{B_m} < U_{B_0}$, are possible. However, if δ is high enough or C or c low enough, then oppositional individuals will be better off. Indeed, on the benefit side, because greens are more numerous, being connected to them give a higher utility to B_m . On the cost side, when C is too high, then B_m is worse off because it is very costly for her to be friend with all the green community. Yet, stability conditions show that, even if these links are costly, she is not interested in severing anyone of these bridge links because the benefits she derives from each of them, due to both the connections and the exposure effects, is larger than the cost of keeping one such link active.

A common feature of all equilibrium networks we have characterized so far is that both communities are fully intracommunity connected. This obviously limits the type of assimilation and identification patterns we could derive. Now, while we still assume very low intra-community costs, i.e. $c < \delta - \delta^2$, we are going to show that, contrary to the standard connections model of Jackson and Wolinsky (1996), in equilibrium, communities can be not fully connected and that both oppositional identities and integration can arise. The following proposition characterizes an extreme form of assimilation.

Proposition 3 *Assume (4). If*

$$C > \frac{n^G + 1}{n^G - 1} [\delta + (n^B - 1) \delta^2 - c] \tag{9}$$

then the network described in Figure 8, where not fully intracommunity communities prevail and where one blue is assimilated and has an oppositional identity while all other blues are separated, is pairwise stable.

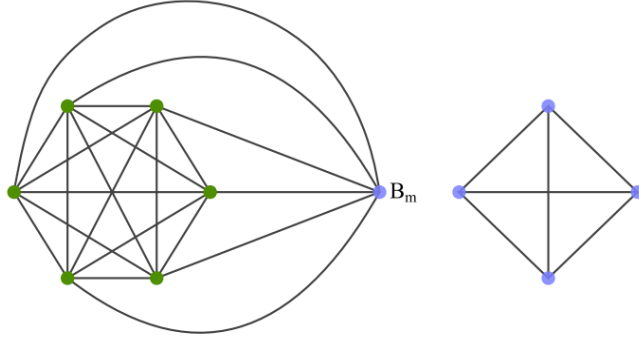


Figure 8. Oppositional identities with non-fully intraconnected communities.

While the condition $\delta - \delta^2 - c > 0$ would always induce fully intraconnected communities in a standard connections model, here, it is not the case since the blue community is fragmented. The logic behind this result is similar to the stability result obtained for the network displayed in Figure 7 (see Proposition 2 (iii)) but in a more extreme form. Indeed, the B_m blue individual now has no connection at all with her own community to minimize her rate of exposure and to eliminate any gap in the linking costs with the green community. In a way, this individual *fully assimilates* to the green community and becomes a green individual according to her choice of social connections. This generates an extreme oppositional identity pattern where a low fraction of blue individuals assimilate with the green community while a majority of blue individuals remain connected only to their community of origin. This is in line with the aggregate conclusions derived from the AddHealth data set we mention in the introduction (see Figure 2). Our proposed mechanism provides a rationale for individual friendship choices. This result highlights the fact that assimilation to the majority culture (Figure 1) makes it difficult for a blue person to interact with her own group. In Section 5.2, we further investigate this case by looking at social norms and sanctions where assimilation to the green culture leads to a rejection from the blue community.

Observe that, in this network (Figure 8), the blue oppositional B_m has always a higher utility than any other non-oppositional blue B_0 since $(\delta - c)n^G > (\delta - c)n^B$. Assimilation with the majority brings access to more social externalities.

The previous result determines under which condition an oppositional identity pattern can emerge in equilibrium. Pushing this logic to the extreme, the next result shows that, if C is sufficiently large, it is even possible that all agents in an economy show an oppositional identity pattern.

Proposition 4 *Assume (4). If C is sufficiently large, the bipartite network in which all green agents are connected to all blue agents, and all blue agents are connected to all green agents is an equilibrium network (Figure 9).*

In the case of a bipartite network each agent is connected only to the other social group and, thus, each agent shows an oppositional identity pattern.

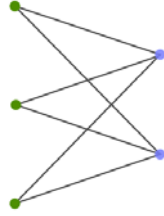


Figure 9. Bipartite Network with $n^W = 3$ and $n^B = 2$.

This network can be sustained in equilibrium because, for an individual of a given type, severing all links to her own community maximizes the positive exposure effect. A link with an agent of same type would be detrimental because, while it would be quite inexpensive in direct terms, it would have a negative counterpart: all links with the agents of the other type would involve a higher cost due to the increase in the fraction of same-type friends, or alternatively, due to the decrease in exposure to the other type. In other words, all green agents are “becoming” blues while all blue agents are “becoming” greens.

3.2 Higher socialization costs

Let us now consider the case when $c > \delta - \delta^2$ so that it becomes more expensive to form links with individuals from the same community. In that range of parameters (i.e. $\delta - \delta^2 < c < \delta$), Jackson and Wolinsky (1996) have shown that, for each community, a star encompassing all individual is always a pairwise stable network.²¹ We thus focus on communities that have a star-shaped form. Of course, since we are dealing with a different cost structure, it is not necessarily true that this result remains valid. However, we are going to present a family of equilibrium networks in which intra-group structure always form a star network.

Proposition 5 *Assume that*

$$\delta - \delta^2 < c < \delta \tag{10}$$

(i) *If*

$$C > \delta + (n^B - 1)\delta^2 - c \tag{11}$$

*then two disconnected star-shaped communities is a pairwise equilibrium network (**complete segregation**). All blues are **separated**.*

²¹Observe that it is not necessarily the unique pairwise stable graph.

(ii) If

$$\delta - \delta^3 - c < C < \delta + (n^B - 1) \delta^2 - c \quad (12)$$

then star-shaped communities connected through their central agents is a pairwise equilibrium network (**partial integration**). Some blues are **separated** and some are **integrated** but none has **oppositional identity**.

(iii) If

$$c > \delta - \delta^3$$

and

$$C < \min \{ \delta + \delta^2 - \delta^4 - \delta^5 - c, 4 [c - (\delta - \delta^3)] \} \quad (13)$$

then star-shaped communities where each peripheral agent has one bridge link with the other peripheral agent whereas stars have no bridge links is a pairwise equilibrium network (**partial integration**). Some blues are **separated** and some are **integrated** but none has **oppositional identity**.

(iv) If

$$C < \delta - \delta^3 - c \quad (14)$$

then star-shaped communities where the centers in both communities are connected to each other and all peripheral agents from both communities are connected to each other is a pairwise equilibrium network (**partial integration**). In that case, **oppositional identities** emerge in equilibrium and all blues are **integrated**.

Figure 10 displays the different cases of Proposition 5 for $n^B = n^G = 3$.

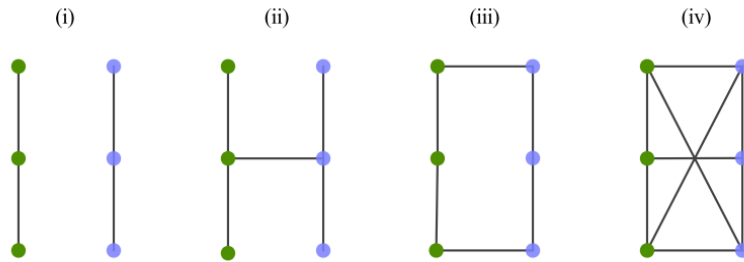


Figure 10. Different equilibrium networks when $\delta - \delta^2 < c < \delta$.

These results are quite intuitive and show how a reduction in C leads to more bridge links and more interactions between communities. Let us explain, for example, why oppositional identities emerge in case (iv), i.e. why some blues have most of their friends who are blues (but are still

integrated) and others have most of their friends who are greens (but are still *integrated*). In case (iv), each peripheral blue (green) individual has one blue (green) friend (the central agent) and $n^G - 1$ ($n^B - 1$) green (blue) friends so that their common same-type friend percentage is $e_i^{\tau(i)} = 1/(n^{\tau(i)})$. This is quite small, especially when the size of the population of each community is large. As a result, each blue (green) peripheral individual displays a high taste for other-type friends, which makes them very attractive. On the contrary, the blue (green) central agent has one green (blue) friend and $n^B - 1$ ($n^G - 1$) blue (green) friends so that $e_i^{\tau(i)} = (n^{\tau(i)} - 1)/n^{\tau(i)}$. This percentage is very close to 1, which make this central agent less attractive for people from the other community. It is now easy to understand why we have oppositional identities. Let us focus on blues. First, peripheral blues do not want to connect to each other because the cost is too high compared to the benefit since $c < \delta - \delta^2$ (they are at distance 2 from each other). Second, peripheral blues do not want to sever a link with one of the $n^G - 1$ peripheral greens because the latter are all very attractive. Finally, peripheral blues do not want to create a link with a central green person because she is not very attractive due to her high intercommunity costs and they can reach him from a peripheral green (distance 2) and obtains δ^2 . This is why peripheral blues have most of their friends who are greens. It is now easy to understand why a blue central individual has most of her friends who are blues. This is due to the fact that she is not attractive (because of her high exposure to her own community) to the peripheral greens.

It is important to observe that this result is *not* due to the size of the communities. It is easy to verify that it still holds if $n^B = n^G = n/2$. More generally, we can see here that there are *reinforcing effects* because once someone from one community is connected to someone from the other community, she becomes more attractive to people from the other community because she costs less in the sense that she is less isolated.

In terms of equilibrium utility, let us study the most interesting case, i.e. (iv). The utility of the peripheral individual (oppositional) is

$$U_P = n^G \delta + (n^B - 1) \delta^2 - c - \left[c + \frac{C}{n^G n^B} \right] (n^G - 1)$$

while that of the center (non-oppositional) individual is:

$$U_C = n^B \delta + (n^G - 1) \delta^2 - (n^B - 1) c - \left[c + \frac{(n^B - 1)(n^G - 1)}{n^G n^B} C \right]$$

We have

$$\begin{aligned} U_P &\stackrel{\geq}{\leq} U_C \\ \Leftrightarrow \frac{(n^G - 1)(n^B - 2)}{n^G n^B} C &\stackrel{\geq}{\leq} (n^G - n^B)(c + \delta^2 - \delta) \end{aligned}$$

As above, this condition is not incompatible with (14) and thus the oppositional individual can have a higher or lower utility than the non-oppositional one depending on the values of C and δ .

4 Social welfare: Integration versus segregation

We now consider some welfare implications of our model. We have previously focused on how decentralized linking decisions can lead to different social network structures. In particular, our model naturally leads to multiple equilibria. Our analysis in Proposition 1, for example, shows that there is a range of parameters in which two extreme outcomes, the complete network (in which all pair of agents, whatever their types, are connected) and a segregated network (in which only connections within communities are established) are both stable networks. The former represents a situation of *social integration* while the latter represents *social segregation*. In terms of efficiency considerations, one may wonder which of the two outcomes is better from a social viewpoint. We shed here some light on this issue by showing the most important source of inefficiencies in our model.

We consider a utilitarian perspective, where social welfare is measured by the unweighted sum of individual utilities. Thus, a network g is *socially preferable* to another network g' whenever the sum of individual utilities in g is higher than the sum of individual utilities in g' , i.e. $\sum_i u_i(g) > \sum_i u_i(g')$.

The following result compares the social welfare of segregated and integrated networks, and states which one is socially preferable.

Proposition 6 *Assume $c < \delta - \delta^2$ and (6). If*

$$n^B (n^G - 1) (n^B - 1) \leq (n - 1)^2 \quad (15)$$

holds, then there exists a threshold \tilde{C} such that for $C \leq \tilde{C}$, integration is efficient whereas when $C \geq \tilde{C}$, segregation is efficient.

This result suggests that, depending on the size of relative social groups, we can not plead for integrated or segregated socialization patterns *a priori*. The possible inefficiency of the integrated network comes from the fact that, for any individual, it is costly to keep all her intercommunity links. Indeed, stability means that it is suboptimal for her to sever one intracommunity link to increase the exposure effect, and therefore she is paying a cost that is proportional to C . When C increases, these costs may overcome the benefits derived from connecting to the other community. Note that this does not contradict stability: when all the rest of her community is connected to the other community, it is optimal for her to also connect to the other community. This is because the exposure rate of any member of the other community, and therefore the cost of directly connecting with each of them, is low precisely because of these intercommunity connections with the rest of the group. Yet, from a collective point of view, each community would be better off in isolation because the aggregate socialization costs are too large when C lies above the threshold \tilde{C} . As a result, the effects of exposure on costs can explain the possible inefficiency of interactions.

This result allows us to link the cost mechanism we explore in this paper with some policy issues. In particular, we can extract some preliminary conclusions on the possible (in)effectiveness of policies that can favor socialization and thus interaction between different communities. Policies that diminish intracommunity socialization costs are not necessarily going to induce more desirable network structures. For example, activities outside the classroom for adolescents or cultural activities at the neighborhood level can favor integrated patterns since they may facilitate interactions among individuals with different identities but the outcome is not going to be socially efficient unless these policies sufficiently decrease the cost of interactions. While the integrated network can be sustained in equilibrium, this equilibrium can be socially undesirable because individuals are exerting an excessive cost to keep their connections with the other community.

5 Extensions

In what follows we discuss two possible extensions of the model. In the first one, we provide a brief discussion of the model where benefits from connections are type-dependent. In the second extension, we discuss how social norms and, in particular, social punishments for deviating from the social behavior of the rest of the community, can also influence assimilation and oppositional identity patterns. The analysis of both situations is far from exhaustive, but in both cases, it gives good hints on how the different effects of the model (the connections and exposure effects, and mutual consent) are affected by these new assumptions.

5.1 Different externalities

We now extend our model by considering different benefits from interacting with others. Basically, if someone (whatever her type) has a link with a green (blue), she obtains a direct benefit of δ_G (δ_B). We also assume the same structure for indirect benefits. For example, if someone is connected to a green who has a blue friend, then she gets $\delta^G + \delta^G \delta^B$. The cost structure is exactly as before and given by (3). The benefit δ^τ can be interpreted in different ways. If, for example, we think of teenagers in a school, then δ^τ could represent the human capital of individual i 's parents so that being friend with someone creates positive externalities in terms of education (see e.g., and Zenou, 2006). If, for example, we think of adults in the labor market, then δ^τ could represent the exchange of job information between two connected individuals (see e.g., Calvó-Armengol and Jackson, 2004 and Ioannides and Loury, 2004). As stated above, strong ties are people from the same community while weak ties are those from the other communities. If greens have a better network than blues, then, as argued by Granovetter (1973, 1974), (green) weak ties are superior to (blue) strong ties for providing support in getting a job because closed networks are limited in providing information about possible jobs. In a close network, everyone knows each other, information is shared and so potential sources of information are quickly shaken down, the network quickly becomes redundant

in terms of access to new information. In contrast, Granovetter stresses the strength of weak ties involving a secondary ring of acquaintances who have contacts with networks outside ego's network and therefore offer new sources of information on job opportunities.

We assume that $\delta^G > \delta^B$ so that there is a higher benefit of interacting with a green than with a blue individual, i.e. the direct externality green individuals exert on others is larger than the one exerted by blue individuals. This can be, for example, a representation of externalities in education between teenagers,²² or differences in human capital and information, such as access to job offers in labor markets.

Note that, compared to the previous setting, the assumption of heterogeneous externalities has no impact on exposure effects. Since the cost of an intercommunity link only depends on the exposure rates, and not on the level of externalities, the exposure effects, which measure changes on the costs of connecting with the other community due to the creation of a new link, are unaffected. Any change in the analysis must thus come from mutual consent and the connections effects. To understand this change and its possible impact on equilibrium networks, we focus our attention on the network displayed in Figure 8 (and analyzed in Proposition 3).

This network was an equilibrium network when $\delta^G = \delta^B = \delta$ and $\delta - \delta^2 - c > 0$. Let us now consider the case when $\delta^G = \delta + \frac{1}{2}\Delta$ and $\delta^B = \delta - \frac{1}{2}\Delta$, where $\Delta = \delta^G - \delta^B > 0$ is measuring the externality gap between the two communities. If Δ is small, then, by continuity, the network in Figure 8 is still an equilibrium network (as long as C is large enough). Mathematically, this is because the conditions that determine the stability of this network are linear inequalities. This linear inequalities are strict,²³ and therefore a small perturbation in the values of δ^G and δ^B from the starting point $\delta^G = \delta^B = \delta$ preserves the order of all these inequalities. In terms of intuition, increasing δ^G has a positive impact of the connections effect for each link the blue individual B_m has with the green community. In other words, each green individual becomes more attractive because the cost of connecting to each of them does not change and externalities increase. As a counterpart, the connections effect is negative for each of the green individuals linked to B_m . Indeed, the blue individual becomes less attractive because, while the cost of keeping the connection is the same, the externalities that each green agent receives from this direct link are smaller. Yet, if Δ is small enough, this negative effect is too small to discourage a green individual to be friend with B_m : overall, this link is still valuable.

Things change when Δ becomes larger. If, in particular, we have that:

$$\delta^B = \delta - \frac{1}{2}\Delta < c \Leftrightarrow 2(\delta - c) < \Delta$$

then each green individual has incentives to sever the link with the B_m individual. The externality the green agent receives from the link is smaller than the cost of keeping it active. The connections

²²See, for example, Neal (2006) and Heckman and LaFontaine (2010), which document and give context to the differences in skills between black and white students.

²³Except for a set of parameters of measure 0.

effect for the green individual is of greater (negative) magnitude, and induces a change on the incentives of the green community. Note that this effect is the opposite for the blue agent B_m : the larger is Δ , the larger is the connections effect with the green community. But, because of mutual consent, these links cannot be sustained in equilibrium.

It is the mixture between changes in connections effects and mutual consent that changes the logic of the model when there are heterogeneous externalities. In general, the connections effects increase for the blue community, but decrease for the green community. This may break mutual consent in situations where mutual consent would be preserved under homogenous externalities. Intuitively, as exemplified in the discussion above, this can make it *more difficult for the blue community to assimilate* to the green community.

5.2 Social norms

Let us now go back to the model with homogeneous externalities δ but modify the cost of socialization choices by taking into account *social norms*. There are studies that illustrate the importance of social sanctions and social norms in ethnic groups.²⁴ Anson (1985) relates the story of Eddie Perry, an African-American youth from Harlem, who graduated with honors from Phillips Exeter Academy and won a full four-year fellowship to Stanford. A close mentor of Eddie explained the psychological tension of coming back home in his own neighborhood: “This kid couldn’t even play basketball. They ridiculed him for that, they ridiculed him for going away to school, they ridiculed him for turning white. I know because he told me they did.” (Anson, 1985, p. 205). In his autobiographical essay, Rodriguez (1982) told us about his own story as a Mexican-American from Sacramento who went to college and for whom English became his dominant language. His (extended) family considered him increasingly alien and as he put it: “*Pocho*, they called me. Sometimes, playfully, teasingly, using the tender diminutive –*mi pochito*. Sometimes not so playfully, mockingly, *Pocho* (Rodriguez (1982, p. 29)).²⁵ These two stories of a black person labeled a white man by his black neighbors and an Hispanic labeled a “gringo” by his extended family are strikingly similar and illustrate the idea of social sanctions and social norms imposed by their own communities.²⁶

In what follows, we propose a simple way of incorporating these forms of social norms and sanctions in our model. The two examples mentioned in the previous paragraph share the same characteristics: an individual of a given community is punished because he is deviating from the social norms imposed by his community. In other words, social punishments increase when there

²⁴See Akerlof (1997) and references therein.

²⁵As Akerlof (1987) noted it, “a Spanish dictionary defines the word ‘pocho’ as an adjective meaning ‘colorless’ or ‘bland’. As a noun it means the Mexican-American who, in becoming an American, forgets his native society.

²⁶See also Stack (1976) for an interesting story of social sanctions/norms imposed by two sisters on their third sister who became middle class. Stack explained how the social distance between them increased, especially clear in the mutual care of their respective children.

are strong differences between own exposure rate and the exposure rate of own friends. We can formalize this idea using a social sanction function that depends on these rates of exposure. The social sanction individual i receives from her own community is a function of her rate of exposure, e_i , and the average rate of exposure of her friends from her own community, which we denote by \bar{e}_i and is given by:

$$\bar{e}_i(g) = \frac{\sum_{j \in N_i^{\tau(i)}(g)} e_j^{\tau(j)}(g)}{n_i^{\tau(i)}(g)}.$$

To understand this formula, consider again the network described in Figure 7 and let us calculate the average rate of exposure of B_m 's friends. We obtain:

$$\bar{e}_{B_m}(g) = \frac{3e_{B_0}^B(g)}{3} = e_{B_0}^B(g) = 1$$

We denote by $s(e_i, \bar{e}_i)$, the social sanction imposed to i . The utility function of individual i is then defined as:

$$u_i(g) = \sum_j \delta^{d(i,j)} - \sum_{j \in N_i(g)} c_{ij} - s(e_i, \bar{e}_i) \quad (16)$$

The utility of individual i includes now three different components: the benefits derived from direct and indirect connections, $\sum_j \delta^{d(i,j)}$, the total cost of forming direct links with both communities, $\sum_{j \in N_i(g)} c_{ij}$, and the social sanction imposed on individual i by her community, $s(e_i, \bar{e}_i)$.

According to our interpretation, there are several properties this social sanction function should satisfy:

- (i) The social sanction is positive, i.e. $s(e_i, \bar{e}_i) > 0$, only when $0 < e_i < \bar{e}_i$ because, in that case, individual i spends more time with the other community than the average of her friends from her own community;
- (ii) The social sanction is equal to zero, i.e. $s(e_i, \bar{e}_i) = 0$, if $e_i \geq \bar{e}_i$ or $e_i = 0$, that is when she spends either more time with her community than the average of her friends or no time at all (in which case, no sanction is possible);
- (iii) In the case when $0 < e_i < \bar{e}_i$, the social sanction imposed on individual i is higher the larger is the difference between e_i and \bar{e}_i , i.e.

$$\frac{\partial s(e_i, \bar{e}_i)}{\partial e_i} \leq 0$$

- (iv) Finally, when $0 < e_i < \bar{e}_i$, the effect of a decrease in individual i 's rate of exposure is stronger when the average rate of exposure of the peer group of individual i in her own community is

larger, that is:

$$\frac{\partial^2 s(e_i, \bar{e}_i)}{\partial \bar{e}_i \partial e_i} \leq 0$$

To fix ideas, consider the following social sanction function:

$$s(e_i, \bar{e}_i) = (e_i - \bar{e}_i)^2 \mathbf{1}_{\{0 < e_i < \bar{e}_i\}}$$

where $\mathbf{1}_{\{0 < e_i < \bar{e}_i\}}$ denotes the indicator function on the set $\{0 < e_i < \bar{e}_i\}$. This function satisfies the four above properties since: (i) $s(e_i, \bar{e}_i) = (e_i - \bar{e}_i)^2 > 0$ when $e_i < \bar{e}_i$; (ii) $s(e_i, \bar{e}_i) = 0$ when $e_i \geq \bar{e}_i$ or $e_i = 0$; (iii) and (iv) when $e_i < \bar{e}_i$, we have:²⁷

$$\frac{\partial s(e_i, \bar{e}_i)}{\partial e_i} = 2(e_i - \bar{e}_i) < 0$$

and

$$\frac{\partial^2 s}{\partial \bar{e}_i \partial e_i} = -2 < 0$$

In this new scenario, the social sanction function amplifies the exposure effects since the social sanction adds an implicit cost for individual i in case her exposure to the other community is higher than the exposure of her peers. The direct benefits and cost of a given link are unaffected compared to the initial formulation of the model, which means that the *connections effects* in the two models coincide. To understand the consequences of this new term in the utility function, let us analyze some of the networks we studied before.

Consider, for example, the integrated and segregated networks described in Proposition 1. The social sanction function $s(e_i, \bar{e}_i)$ facilitates the stability of segregated networks. Indeed, when considering creating a link with the other community, for each individual, independently of her type, the connections effect is unaffected because the externalities and the direct cost of building a link are the same as in the initial model. The exposure effect is, however, magnified because creating a bridge link will decrease the exposure rate of each individual involved in the link, which will be below that of the rest of the community. As a result, because of the social sanction, the incentives to create an intercommunity link are lowered and the stability of segregated networks is preserved.

In the analysis of the integrated networks, the social sanction plays a minor role. The social sanction increases the incentives of keeping links with own community since severing one of these links would reduce the exposure rate of this individual below the average exposure rate of the rest of her community, implying a positive social sanction. Given that the connections and exposure

²⁷Economists have modelled *conformity* in a similar way by adding a term $-(e_i - \bar{e}_i)^2$ to the utility function, where e_i and \bar{e}_i is the effort of individual i and the average effort of i 's peers. In that case, each individual i loses utility $(e_i - \bar{e}_i)^2$ from failing to conform to her peers (see, among others, Akerlof, 1980; Bernheim, 1994; Kandel and Lazear, 1992; Akerlof, 1997; Fershtman and Weiss, 1998; Patacchini and Zenou, 2011). Our formulation is slightly different since the social cost is only paid when $e_i < \bar{e}_i$.

effects are the same as in Section 3, this implies that the incentives of keeping an intracommunity link are higher in this case. On the other hand, when individual i wants to sever a link with someone from the other community, her incentives are unaffected. Indeed, the connections and exposure effect are exactly the same and the social sanction is zero since, by severing such a link, her exposure rate becomes higher than that of the rest of her community.

This analysis should make apparent that, in general, social sanctions relax the incentive constraints of individuals facing situations where the relation between communities is very homogeneous across individuals within the same community. The analysis is quite different when we consider networks where individuals from one community show very asymmetric positions such as, for example, the blue individual B_m in the network displayed in Figure 7. In that case, the differences in terms of exposure rates are large and the social sanction from deviating from the community's norm is high. This makes it harder to keep each intercommunity link active, as it is the case in Figure 7 since severing one of these links would mitigate the strength of the social sanction. The incentives to socialize with the other community are now very different for B_m and the network described in Figure 7 is probably not stable (the stability would depend on the particular functional form considered for $s(e_i, \bar{e}_i)$). In this asymmetric network, the social sanction function strengthens her incentives constraints of socializing with the other community instead of relaxing them. There is, however, one caveat. The social sanction function $s(e_i, \bar{e}_i)$ has a discontinuity at $e_i = 0$. If we consider the network displayed in Figure 8, the social sanction relaxes the stability constraints because the blue individual B_m has no connection with her community of origin and therefore does not suffer from any type of social sanction from her community. The incentives to stay *disconnected* with her community of origin are strengthened (because the connections and exposure effects are the same while the social sanction would be very high if building a connection with the blue community) and the stability constraints for the links with the green community are unaffected because the social sanction would remain equal to 0 in the case of severing any of them. The discontinuity of $s(e_i, \bar{e}_i)$ plays a leading role in this argument.

In some sense, Eddie Perry and Richard Rodriguez (mentioned above) have both chosen to have a very low exposure to their own community (i.e. low e_i) and are paying a very high price for it when interacting with people from their community of origin.

6 Conclusion

In this paper, we consider social networks as the main building blocks for individual identity formation. This is a complementary view from that developed in other research (as, for example, in Akerlof and Kranton, 2000), where identities are sometimes interpreted as a direct choice and where it is precisely this unidimensional choice that determines socioeconomic outcomes. The choice of direct network interactions is, instead, necessarily a multidimensional and complex decision. In our case, these decentralized linking decisions are the channel determining each individual's social

capital. We have modeled these decisions through a precise network structure that shapes social interactions and the exposure and assimilation of others' differences.

Identification patterns are important for individual and collective outcomes. Using Swedish data and focusing on the two-dimensional aspect of identity defined in Figure 1, Nekby and Rödin (2009) show that what matters for labor market outcomes is the strength of identification with the majority culture regardless of the strength of own ethnic identity. In other words, having a strong ethnic identity is not necessarily negative for the labor market if it is not associated with a rejection of the majority culture values. Using the same bidimensional measure of identity, Zimmermann et al. (2007), Constant and Zimmermann (2008), Constant et al. (2009) find that, for Germany, human capital acquired in origin countries lead to lower identification with the majority culture while education acquired post-migration in the host country does not affect attachment to the majority culture. Battu and Zenou (2010) find similar results for the UK while Bisin et al. (2011), studying different European countries, show that there is a penalty in the labor market for minorities with a strong identity.

We believe that our model points to an important and still understudied issue in the literature on economics and identity. In particular, our analysis has been able to mimic some characteristics of different real-world networks, such as the rise of oppositional identity patterns. While the model is quite stylized, we believe it is helpful because it provides a neat taxonomy of the strategic effects that are at the root of individual decisions for building connections. The extensions considered in Section 5 show how these effects would be perturbed when new elements are introduced into the model.

In what follows, we suggest three avenues for future research that seem particularly promising.

Initial exogenous categories, reflected in our model by the initial assignment of one of the two possible types, are reasonable in some setups. For example, family endows each individual with some cultural traits, such as inherited language. Yet, in other setups, we expect that both the initial identity (type) and direct connections (network) are an individual choice. This can be the case, for example, for adolescent behavior in the classroom.²⁸ It would be interesting to encompass in a unified framework both dimensions of choices and to study the interplay of both the individual and social dimensions in the determination of identity. Presumably, in this richer framework, there might be complementarities in the final strategies of each individual in both dimensions: the choice in one dimension correlates and amplifies the choice in the other dimension.

From a more technical perspective, it would also be worth studying possible refinements of our equilibrium concept that could help providing more precise results and a more exhaustive characterization of the set of equilibrium networks. This is going to increase the already important combinatorial complexity in the analysis, which already deprives us from obtaining a full charac-

²⁸See, for example, Coleman (1961), where a taxonomy of identities that adolescents adopt in US high-schools is provided, and Akerlof and Kranton (2002), for an economic analysis of identity in schooling.

terization of pairwise stable networks.

Finally, we have not deepened other important consequence of network structure, such as segregation and inequality. A recent work by Echenique and Fryer (2007) has introduced a new measure of individual segregation rooted at the social network level. This measure could be used in our setup to analyze the segregation patterns emerging from decentralized network formation. Kets et al. (2011) have also proposed an interesting model exploring how the structure of a social network constrains the level of inequality that can be sustained among its members. In their model, what influences inequality is the ability of players to form viable coalitions given an exogenous social network. It would be interesting to relate network formation and identity considerations to these relevant issues.

References

- [1] Abrams D. and M. Hogg (1999), *Social Identity and Social Cognition*, Oxford: Blackwell Publishers.
- [2] Ainsworth-Darnell, J.W. and D.B. Downey (1998), “Assessing the oppositional culture explanation for racial/ethnic differences in school performance,” *American Sociological Review* 63, 536-553.
- [3] Akerlof, G.A. (1980), “A theory of social custom of which unemployment may be one consequence,” *Quarterly Journal of Economics* 94, 749-775.
- [4] Akerlof, G.A. (1997), “Social distance and social decisions,” *Econometrica* 65, 10005-1027.
- [5] Akerlof, G.A. and R.E. Kranton (2000), “Economics and identity”, *Quarterly Journal of Economics* 115, 715-753.
- [6] Akerlof, G.A. and R.E. Kranton (2002), “Identity and schooling: some lessons for the economics of education”, *Journal of Economic Literature* 40, 1167-1201.
- [7] Akerlof, G.A. and R.E. Kranton (2010), *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*, Princeton: Princeton University Press.
- [8] Anson, R.S. (1985), *Best Intentions: The Education and Killing of Edmund Perry*, New York: Random House.
- [9] Austen-Smith D. and R.D. Fryer, Jr (2005), “An economic analysis of ‘acting white’,” *Quarterly Journal of Economics* 120, 551-583.
- [10] Battu, H. McDonald, M. and Y. Zenou (2007), “Oppositional identities and the labor market,” *Journal of Population Economics* 20, 643-667.

- [11] Battu, H. and Y. Zenou (2010), “Oppositional identities and employment for ethnic minorities. Evidence for England,” *Economic Journal* 120, F52-F71.
- [12] Ben-Ner, A., McCall, B.P., Stephanec, M. and H. Wang (2009), “Identity and in-group/out-group differentiation in work and giving behaviors: Experimental evidence,” *Journal of Economic Behavior and Organization* 72, 153-170.
- [13] Benabou, R. (1993), “Workings of a city: Location, education, and production,” *Quarterly Journal of Economics* 108, 619-652.
- [14] Bernheim, B.D. (1994), “A theory of conformity,” *Journal of Political Economy* 102, 841-877.
- [15] Berry, J.W. (1997), “Immigration, acculturation and adaptation,” *Applied Psychology: An International Review* 46, 5-34.
- [16] Bisin, A., Patacchini, E., Verdier, T. and Y. Zenou (2010), “Formation and persistence of oppositional identities,” Unpublished manuscript, Stockholm University.
- [17] Bisin, A., Patacchini, E., Verdier, T. and Y. Zenou (2011), “Ethnic identity and labor-market outcomes of immigrants in Europe,” *Economic Policy* 26, 57-92.
- [18] Buhai, S. and M. van der Leij (2008), “A social network analysis of occupational segregation,” Unpublished manuscript, Tinbergen Institute.
- [19] Bramoullé, Y. and B.W. Rogers (2009), “Diversity and popularity in social networks,” Unpublished manuscript, Northwestern University.
- [20] Braugh, J. (1983), *Black Street Speech: Its History, Structure and Survival*, Austin: University of Texas Press.
- [21] Cabrales, A., Calvó-Armengol, A. and Y. Zenou (2011), “Social interactions and spillovers,” *Games and Economic Behavior*, forthcoming.
- [22] Calvó-Armengol, A. and M.O. Jackson (2004), “The effects of social networks on employment and inequality,” *American Economic Review* 94, 426-454.
- [23] Camargo, B., Stinebrickner, R. and T. Stinebrickner (2010), “Interracial friendships in college,” NBER Working paper No. 15970.
- [24] Coleman, J.S. (1958), “Relational analysis: the study of social organizations with survey methods,” *Human Organization* 17, 28-36.
- [25] Coleman, J.S. (1961), *The Adolescent Society: The Social Life of the Teenager and Its Impact on Education*, New York: The Free Press.

- [26] Constant, A., Gataullina, L. and K.F. Zimmermann (2009), “Ethnosizing immigrants,” *Journal of Economic Behavior and Organization* 69, 274-287.
- [27] Constant, A. and K.F. Zimmermann (2008), “Measuring ethnic identity and its impact on economic behavior,” *Journal of the European Economic Association* 6, 424-433.
- [28] Currarini, S., Jackson, M.O., and P. Pin (2009), “An economic model of friendship: Homophily, minorities, and segregation,” *Econometrica* 77, 1003-1045.
- [29] Currarini, S., Jackson, M.O., and P. Pin (2009), “Identifying the roles of race-based choice and chance in high school friendship network formation,” *Proceedings of the National Academy of Sciences* 107, 4857-4861.
- [30] Cutler, D., Glaeser, E.L. and Vigdor, J.L. (1999), “The rise and decline of the American ghetto,” *Journal of Political Economy* 107, 455-506.
- [31] Delpit, L. (1995), *Other People’s Children: Cultural Conflict in the Classroom*, New York: The Free Press.
- [32] Echenique, F. and R.D. Fryer, Jr (2007), “A measure of segregation based on social interactions,” *Quarterly Journal of Economics* 122, 441-485.
- [33] Eguia, J.X. (2010), “A theory of discrimination and assimilation,” Unpublished paper, New York University.
- [34] Fershtman, C. and Y. Weiss (1998), “Social rewards, externalities and stable preferences,” *Journal of Public Economics* 70, 53-73.
- [35] Fordham S. and J.U. Ogbu (1986), “Black student’s school success: Coping with the burden of acting White,” *The Urban Review* 18, 176-206.
- [36] Fryer, R.G. Jr. and P. Torelli (2010), “An empirical analysis of ‘acting white’,” *Journal of Public Economics* 94, 380-396.
- [37] Golub, B. and Jackson M. (2008), “How homophily affects diffusion and learning in networks,” Unpublished manuscript, Stanford University.
- [38] Goyal, S. (2007), *Connections: An Introduction to the Economics of Networks*, Princeton: Princeton University Press.
- [39] Goyal, S. and F. Vega-Redondo (2007), “Structural holes in social networks,” *Journal of Economic Theory* 137,460-492.
- [40] Granovetter, M.S. (1973), “The strength of weak ties,” *American Journal of Sociology* 78, 1360-1380.

- [41] Granovetter, M.S. (1974), *Getting a Job: A Study of Contacts and Careers*, Cambridge, MA: Harvard University Press.
- [42] Heckman, J.J. and P.A. LaFontaine (2010), “The American high school graduation rate: Trends and levels,” *Review of Economics and Statistics* 92, 244-262.
- [43] Iijima, R. and Y. Kamada (2010), “Social distance and network structures,” Unpublished manuscript, Harvard University.
- [44] Ioannides, Y.M. and D.L. Loury (2004), “Job information networks, neighborhood effects, and inequality,” *Journal of Economic Literature* 42, 1056-1093.
- [45] Jackson, M.O. (2008), *Social and Economic Networks*, Princeton: Princeton University Press.
- [46] Jackson, M.O. and B.W. Rogers (2005), “The economics of small worlds,” *Journal of the European Economic Association* 3, 617-627.
- [47] Jackson, M.O. and A. Wolinsky (1996), “A strategic model of social and economic networks,” *Journal of Economic Theory* 71, 44-74.
- [48] Johnson, C. and R.P. Gilles (2000), “Spatial social networks,” *Review of Economic Design* 5, 273-299.
- [49] Kandel, E. and, E.P. Lazear (1992), “Peer pressure and partnerships,” *Journal of Political Economy* 100, 801-817.
- [50] Kets, W., Iyengar, G., Sethi, R. and S. Bowles (2011), “Inequality and network structure,” *Games and Economic Behavior*, forthcoming.
- [51] Kirman, A. and M. Teschl (2004), “On the emergence of economic identity,” *Revue de Philosophie Economique* 9, 59-86.
- [52] Labov, W. (1972), *Language in the Inner City: Studies in the Black English Vernacular*, Philadelphia: University of Pennsylvania Press.
- [53] Labov, W. and W.A. Harris (1986), “De facto segregation of black and white vernaculars, in D. Sankoff (Ed.), *Current Issues in Linguistic Theory 53: Diversity and Diachrony*, Amsterdam: John Benjamins Publishing, pp. 1-24.
- [54] Lang, K. (2007), *Poverty and Discrimination*, Princeton: Princeton University Press.
- [55] Lemanski, C. (2007), “The impact of residential desegregation on social integration: Evidence from a South African neighbourhood,” *Geoforum* 37, 417-435.

- [56] Marmaros, D. and B. Sacerdote (2006), “How do friendships form?” *Quarterly Journal of Economics* 121, 79-119.
- [57] Mayer, A. and S.L. Puller (2008), “The old boy (and girl) network: Social network formation on university campuses,” *Journal of Public Economics* 92, 329-347.
- [58] Mele, A. (2010), “A structural model of segregation in social networks,” Unpublished manuscript, University of Illinois at Urbana-Champaign.
- [59] Myerson, R.B. (1991), *Game Theory: Analysis of Conflict*, Harvard: Harvard University Press.
- [60] Mobius, M. (2007), “The Formation of ghettos as a local interaction phenomenon,” Unpublished manuscript, Harvard University.
- [61] Neal, D. (2006), “Why has black-white skill convergence stopped?” In: E. Hanushek and F. Welch (Eds.), *Handbook of Economics of Education*, Amsterdam: Elsevier Science.
- [62] Nekby, L. and M. Rödin (2009), “Acculturation identity and employment among second and middle generation immigrants,” *Journal of Economic Psychology* 31, 35-50.
- [63] Ogbu, J.U. (2003), *Black American Students in an Affluent Suburb: A Study of Academic Disengagement*, New Jersey: Lawrence Erlbaum.
- [64] Patacchini, E. and Y. Zenou (2006), “Racial identity and education,” CEPR Discussion Paper No. 5607.
- [65] Patacchini, E. and Y. Zenou (2011), “Juvenile delinquency and conformism,” *Journal of Law, Economics, and Organization*, forthcoming.
- [66] Phinney, J.S. (1990), “Ethnic identity in adolescents and adults: Review of research,” *Psychological Bulletin* 108, 499-514.
- [67] Rodriguez, R. (1982), *Hunger of Memory: The Education of Richard Rodriguez*, New York: Bantam.
- [68] Ryder A.G., Alden, L.E. and D.L. Paulhus (2000), “Is acculturation unidimensional or bidimensional? A head-to-head comparison in the prediction of personality, self-identity, and adjustment,” *Journal of Personality and Social Psychology* 79, 49-65.
- [69] Schelling, T.C. (1971), “Dynamic models of segregation,” *Journal of Mathematical Sociology* 1, 143-186.
- [70] Selod, H. and Y. Zenou (2006), “City-structure, job search, and labour discrimination. Theory and policy implications,” *Economic Journal* 116, 1057-1087.

- [71] Stack, C.B. (1976), *All Our Kin: Strategies for Survival in a Black Community*, New York: Harper & Row.
- [72] Wilson, W. J. (1987), *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*, Chicago: University of Chicago Press.
- [73] Zhang, J. (2004), “Residential segregation in an all-integrationist world,” *Journal of Economic Behavior and Organization* 54, 533-550.
- [74] Zimmermann, L., Zimmermann, K.F. and A. Constant (2007), “Ethnic self-identification of first-generation immigrants,” *International Migration Review* 41, 769–781.

APPENDIX A: PROOFS

Proof of Proposition 1.

(i) *Complete integration* between communities: There is *no* gain in utility for a green person to sever a link with a blue person, who is necessarily connected to the rest of the blue community, if:

$$\begin{aligned} & \delta - \delta^2 - c - \left(\frac{n^G - 1}{n - 2}\right) \left(\frac{n^B - 1}{n - 2}\right) C \\ & + (n^B - 1) \left[\left(\frac{n^B - 1}{n - 2}\right) \left(\frac{n^G - 1}{n - 3}\right) - \left(\frac{n^B - 1}{n - 2}\right) \left(\frac{n^G - 1}{n - 2}\right) \right] C \geq 0 \end{aligned} \quad (17)$$

The first term $\delta - \delta^2$ are the benefits derived from externalities of having a direct connection instead of an indirect connection with this blue person. The second term, $-c - \left(\frac{n^G - 1}{n - 2}\right) \left(\frac{n^B - 1}{n - 2}\right) C$,²⁹ is the cost of forming the link with this blue person. The sum of these two first terms is what we refer in the main text as the *connections effect*. The third term,

$$(n^B - 1) \left[\left(\frac{n^B - 1}{n - 2}\right) \left(\frac{n^G - 1}{n - 3}\right) - \left(\frac{n^B - 1}{n - 2}\right) \left(\frac{n^G - 1}{n - 2}\right) \right] C$$

is the indirect benefit derived from the diminishing costs of maintaining a link with a blue person, once this new link is formed. Before forming the new link, the proportion of green friends among all green person's friends is $\frac{n^G - 1}{n - 3}$. Once the new link is created, this proportion diminishes to $\frac{n^G - 1}{n - 2}$, and this implies a decrease in the cost of maintaining the link with the $n^B - 1$ blue persons from $c + \left(\frac{n^B - 1}{n - 2}\right) \left(\frac{n^G - 1}{n - 3}\right) C$ to $c + \left(\frac{n^B - 1}{n - 2}\right) \left(\frac{n^G - 1}{n - 2}\right) C$. The third term in (17) accounts for this difference in costs and was referred to as the *exposure effect* in the main text.

The inequality (17) is equivalent to

$$\delta - \delta^2 - c \geq \left(\frac{n^B - 1}{n - 2}\right) (n^G - 1) \left[\frac{n^B - n + 2}{(n - 3)(n - 2)} \right] C$$

which is always true since by assumption $n \geq n^B + 2$, which implies that the term in the right hand side of this last expression is negative, and $\delta - \delta^2 - c > 0$ by assumption. Because of symmetry, a similar argument can be used to guarantee that *no* gain in utility can be found for a blue person that severs a link with a green person.

We have now to check that there is *no* gain in utility for a green person to sever a link with another green person. The relevant inequality to make sure that a green individual does not have such incentives is:

$$\delta - \delta^2 - c \quad (18)$$

²⁹The $n - 2$ in the denominators come from the fact we are including the individuals involved in the link in the computation of exposure rates.

$$+n^B \left[\left(\frac{n^B - 1}{n - 2} \right) \left(\frac{n^G - 2}{n - 3} \right) - \left(\frac{n^B - 1}{n - 2} \right) \left(\frac{n^G - 1}{n - 2} \right) \right] C \geq 0$$

The strength of the connections effect is now equal to $\delta - \delta^2 - c$ because we are considering an intracommunity link and, hence, its cost is equal to c . When a green agents severs a link with another green individual, she is reducing her exposure to her own community and this diminishes the cost of each of the links with the other community. This exposure effect is measured by the term in the second line of (18). This term is negative, meaning that severing a link decreases the intracommunity costs. Inequality (18) is equivalent to

$$C \leq \frac{(n - 2)^2 (n - 3)}{n^B (n^B - 1)^2} (\delta - \delta^2 - c)$$

An equivalent argument shows that a blue individual does not want to sever a link with another blue individual if and only if

$$C \leq \frac{(n - 2)^2 (n - 3)}{n^G (n^G - 1)^2} (\delta - \delta^2 - c)$$

Since $n^G \geq n^B$, this last inequality is the condition that ensures that no individual has incentives to sever a link with her own community.

(ii) Let us show that *complete segregation* between communities is an equilibrium network. There is *no* gain in utility for a green person to establish a link with a blue person, who is necessarily connected to the rest of the blue community, if:

$$\delta + (n^B - 1) \delta^2 - c < C$$

Similarly, there is *no* gain in utility for a blue person to connect to a green individual, who is necessarily connected to the rest of the green community, if:

$$\delta + (n^G - 1) \delta^2 - c < C$$

Since $n^G \geq n^B$, and because mutual consent is necessary, then condition (6) guarantees that there is complete segregation. ■

Proof of Proposition 2

(i) Let us denote by G_1 (resp. B_1) the unique green (resp. unique blue) agent involved in the bridge link. Agent G_1 has incentives to form a link with B_1 iff

$$\delta + (n^B - 1) \delta^2 - c > C \tag{19}$$

Similarly, agent B_1 has incentives to form the link with G_1 iff

$$\delta + (n^G - 1) \delta^2 - c > C \tag{20}$$

Since $n^G \geq n^B$, the first condition is more restrictive than the second. Mutual consent in link formation imposes that both conditions have to be satisfied at the same time, hence (19) is a requirement for the network to be pairwise stable.

Under assumption (4), we know that both communities are fully intraconnected. We have to check that no other pair of agents of distinct types other than G_1 and B_1 have incentives to form a link.

The green agent G_1 does not have incentives to form a link with a blue one who is not B_1 iff

$$\delta - \delta^2 - \left[c + \frac{(n^G - 1)}{n^G} C \right] + \left[1 - \frac{(n^G - 1)}{n^G} \right] C < 0$$

which is equivalent to:

$$\left(\frac{n^G}{n^G - 2} \right) (\delta - \delta^2 - c) < C \quad (21)$$

The blue agent B_1 does not have incentives to form a link with a green agent who is not G_1 iff

$$\delta - \delta^2 - \left[c + \left(\frac{n^B - 1}{n^B} \right) C \right] + \left[1 - \left(\frac{n^B - 1}{n^B} \right) \right] C < 0$$

which is equivalent to:

$$\left(\frac{n^B}{n^B - 2} \right) (\delta - \delta^2 - c) < C \quad (22)$$

Because of mutual consent and since $n^G \geq n^B$ only condition (21) is required.

Any other green agent different than G_1 does not have incentives to form a link with B_1 iff

$$\left(\frac{n^B}{n^B - 1} \right) [(1 - \delta) (\delta + (n^B - 1) \delta^2) - c] < C$$

which is equivalent to

$$\left(\frac{n^B}{n^B - 1} \right) [\delta + (n^B - 2) \delta^2 - (n^B - 1) \delta^3 - c] < C \quad (23)$$

Because of symmetry, any other blue agent different than B_1 does not have incentives to form a link with G_1 iff

$$\left(\frac{n^G}{n^G - 1} \right) [\delta + (n^G - 2) \delta^2 - (n^G - 1) \delta^3 - c] < C \quad (24)$$

Finally, any green agent other than G_1 does not have incentives to form a link with a blue other than B_1 iff

$$\delta - \delta^3 + (n^B - 2) (\delta^2 - \delta^3) - c < C$$

which is equivalent to

$$\delta + (n^B - 2) \delta^2 - (n^B - 1) \delta^3 - c < C \quad (25)$$

Any blue agent other than B_1 does not have incentives to form a link with a green different than G_1 iff

$$\delta - \delta^3 + (n^G - 2) (\delta^2 - \delta^3) - c < C$$

which is equivalent to

$$\delta + (n^G - 2) \delta^2 - (n^G - 1) \delta^3 - c < C \quad (26)$$

Because of mutual consent, only one of the conditions among (25) and (26) has to hold. Observe that if either (23) or (24) holds, then either (25) or (26) hold too.

Note that since $\delta > \delta^2 + c$, no green or blue individual different than G_1 or B_1 has incentives to sever a link with her own community. In this case, there is no exposure effect to consider because such individuals have no ties with the other community.

Gathering everything together the required conditions for the network to be pairwise stable are given by (19), (21), (23) and (24).

(ii) In this network, a green agent with a bridge link does not have incentives to sever it iff

$$\delta - \delta^3 - c > C \quad (27)$$

Because of symmetry, this same condition ensures that a blue agent with a bridge link does not have incentives to sever it.

A green agent that has already one bridge link does not have incentives to build a new one iff

$$\delta - \delta^2 - c + \left[1 - \left(\frac{n^G - 1}{n^G} \right) \right] C < \left(\frac{n^G - 1}{n^G} \right) \left(\frac{n^B - 1}{n^B} \right) C$$

which is equivalent to

$$\frac{n^G n^B}{(n^G - 1)(n^B - 1) - n^B} (\delta - \delta^2 - c) < C \quad (28)$$

Similarly, a blue agent does not have incentives to build a new bridge link with a green that has already a bridge link iff

$$\left[\frac{n^G n^B}{(n^G - 1)(n^B - 1) - n^G} \right] (\delta - \delta^2 - c) < C \quad (29)$$

Because of mutual consent, only (28) or (29) is needed. Since the first of these conditions is less restrictive, it suffices to ensure that this type of link is not formed.

Furthermore, a blue agent with a bridge link does not have incentives to build a link with a green that does not have a bridge link iff

$$\delta - \delta^2 - c - \left(\frac{n^B - 1}{n^B} \right) C + \left(1 - \frac{n^B - 1}{n^B} \right) C < 0$$

which is equivalent to

$$\left(\frac{n^B}{n^B - 2} \right) (\delta - \delta^2 - c) < C \quad (30)$$

A green that is not in a bridge does not have incentives to create a link with a blue agent iff

$$\left(\frac{n^B}{n^B - 1}\right) (\delta - \delta^2 - c) < C \quad (31)$$

Because of mutual consent, only (30) or (31) is needed. since the second of these conditions is less restrictive, it suffices to ensure that this type of link is not formed.

Since

$$\frac{n^G n^B}{(n^G - 1)(n^B - 1) - n^G} > \frac{n^B}{n^B - 1} \Leftrightarrow (n^B)^2 + n^G n^B > n^B \Leftrightarrow n^B + n^G > 1$$

condition (28) implies (30).

Finally, note that, as in the previous section, the green individuals that do not have any bridge link with the other community do not have incentives to sever any of their links with their community. Similarly, the rest of green and blue individuals, which have one bridge link, have no incentives to sever an intracommunity link because this has no effect in the cost of their bridge link. This is because their exposure rate in the computation of the cost of the bridge link, which is equal to 1 for both sides of the link, is not affected if intracommunity link is severed. This is a general rule that we use in subsequent proofs: when an individual has at most one bridge link, severing or creating intracommunity links generates no exposure effect.

Hence, gathering everything together we obtain that the two required conditions are (27) and (28).

(iii) We call ‘‘oppositional blue’’ and denote her by B_m , the blue agent that has a bridge link with each of the members of the green community. The oppositional blue individual B_m does not want to sever any of her bridge links iff

$$\begin{aligned} \delta - \delta^2 - c - \left(\frac{n^B - 1}{n - 2}\right) C + (n^G - 1) \left[\left(\frac{n^B - 1}{n - 3}\right) - \left(\frac{n^B - 1}{n - 2}\right) \right] C > 0 \\ \Leftrightarrow \delta - \delta^2 - c - \left(\frac{n^B - 1}{n - 2}\right) C + \frac{(n^G - 1)(n^B - 1)}{(n - 3)(n - 2)} C > 0 \end{aligned}$$

which is equivalent to

$$\frac{(n - 2)(n - 3)}{(n^B - 1)(n^B - 2)} (\delta - \delta^2 - c) > C \quad (32)$$

A green agent does not want to sever his bridge link with the oppositional blue B_m iff

$$\delta - \delta^2 + (n^B - 1) (\delta^2 - \delta^3) - c - \left(\frac{n^B - 1}{n - 2}\right) C > 0$$

which is equivalent to

$$\left(\frac{n - 2}{n^B - 1}\right) [(1 - \delta) (\delta + (n^B - 1) \delta^2) - c] > C \quad (33)$$

Any of the non-oppositional blues, denoted by B_0 , does not have incentives to directly connect with a green agent iff

$$\delta - \delta^2 - c - \left(\frac{n^G - 1}{n^G}\right) C < 0$$

which is equivalent to

$$\left(\frac{n^G}{n^G - 1}\right) [\delta - \delta^2 - c] < C \quad (34)$$

A green agent does not have incentives to connect with a non-oppositional blue B_0 iff

$$\begin{aligned} \delta - \delta^2 - c - \left(\frac{n^G - 1}{n^G}\right) C + \left[\left(\frac{n^B - 1}{n - 2}\right) - \left(\frac{n^G - 1}{n^G}\right) \left(\frac{n^B - 1}{n - 2}\right)\right] C < 0 \\ \Leftrightarrow \delta - \delta^2 - c - \left(\frac{n^G - 1}{n^G}\right) C + \left(\frac{n^B - 1}{n - 2}\right) \left(\frac{1}{n^G}\right) C < 0 \end{aligned}$$

which is equivalent to

$$\frac{\delta - \delta^2 - c}{\left(\frac{n^G - 1}{n^G}\right) - \left(\frac{n^B - 1}{n - 2}\right) \left(\frac{1}{n^G}\right)} < C \quad (35)$$

The conditions (32) and (33) have to hold. Because of mutual consent, only one of the conditions (34) and (35) is required. Condition (34) is less restrictive than the last one, and hence, the set of required conditions are (32), (33) and (34), that need to hold at the same time. ■

Proof of Proposition 3

Consider the network described in Figure 8. There are n^G individuals who are all connected with each other. There is one blue B_m who is connected to all greens and is not connected to any other blue B_0 . All the other $n^B - 1$ blues are fully connected with each other.

The blue individual B_m does not want to create a link with a blue individual B_0 iff

$$n^G \delta - n^G c - n^G \left(0 \times \frac{n^G - 1}{n^G}\right) C > (n^G + 1) \delta + (n^B - 1) \delta^2 - c - n^G c - n^G \left(\frac{1}{n^G + 1} \frac{n^G - 1}{n^G}\right) C$$

which is equivalent to

$$C > \frac{[\delta + (n^B - 1) \delta^2 - c] (n^G + 1)}{n^G - 1}$$

The blue individual B_m does not want to sever a link with a green individual iff

$$n^G \delta - n^G c > (n^G - 1) \delta + \delta^2 - (n^G - 1) c$$

which is equivalent to

$$c < \delta - \delta^2$$

This is always true because of assumption (4).

Observe that, because of (4), none of the blues B_0 would like to sever a link with a B_0 . Because of mutual consent, when condition (9) holds, they cannot have a link with B_m because B_m does not want to.

A green will not sever a link with B_m iff

$$\delta - c > \delta^2$$

which is always true because of (4).

Finally, because of (4), it follows that a green does not want to sever a link with another green. This is again, as in the case of the Proof of Proposition 2 *ii*) because for green agents there is no exposure effect to consider since in the network configuration we are analyzing they have only one bridge link with the blue community.

Therefore, condition (9) is enough to guarantee that the network described by Figure 8 is pairwise stable. ■

Proof of Proposition 4

A green agent does not have incentives to sever a link with a blue agent whenever

$$\delta - \delta^3 - c \geq 0,$$

which is immediately satisfied when $c < \delta - \delta^2$.

On the other hand, a green agent does not have incentives to create a link with another green agent if

$$\delta - \delta^2 - c - n^B \left(\frac{1}{n^B - 1} C \right) < 0$$

This condition is satisfied if C is high enough.

A similar argument holds for a blue agent. ■

Proof of Proposition 5

(*i*) The center in the star formed by the green community (that we call the “green center”) does not have incentives to build a link with the center in the star formed by the blue community (the blue center) iff

$$\delta + (n^B - 1) \delta^2 - c < C$$

Similarly, the blue center does not have incentives to build a link with the green center iff

$$\delta + (n^G - 1) \delta^2 - c < C$$

Because of mutual consent, only the first one, that coincides with (11), needs to hold

If the centers have no incentive to connect with each other, *a fortiori* no individual of one community has incentives to connect with an individual of the other community.

(ii) Firstly, if

$$\delta + (n^B - 1) \delta^2 - c > C$$

then neither the green center nor the blue center has incentives to sever the bridge link that connects them.

Furthermore, none of the centers has incentives to sever a link with her own community because they have just one bridge link with the other community, and we can once more apply the result that there is no *exposure effect* in this case.

The green center does not have incentives to connect with a peripheral agent of the blue community iff

$$\left(\frac{n^G}{n^G - 2} \right) (\delta - \delta^2 - c) < C$$

Observe that this is satisfied by assumption.

Similarly, the blue center does not have incentives to connect with a peripheral agent of the green community iff

$$\left(\frac{n^B}{n^B - 2} \right) (\delta - \delta^2 - c) < C$$

Again, this condition is trivially satisfied.

Since these last two conditions ensure that none of the centers have incentives to form a link with the periphery of the other community, and since mutual consent is necessary for link formation, we don't have to check for the conditions that ensure that a peripheral agent does not have incentives to connect with the center of the other community.

A green peripheral agent does not have incentives to connect with a blue peripheral iff

$$(1 - \delta^2) \delta - c < C$$

Because of symmetry, this same condition ensures that a blue peripheral agent does not have incentives to connect with a green peripheral agent. Hence, this last condition, jointly with $\delta + (n^B - 1) \delta^2 - c > C$, ensure that the network analyzed is stable.

(iii) A green peripheral agent does not have incentives to sever his bridge link with a blue peripheral iff

$$\delta - \delta^5 + \delta^2 - \delta^4 - c > C \tag{36}$$

This same condition ensures that a blue peripheral agent does not have incentives to sever his bridge link with a green peripheral agent.

A green peripheral agent does not have incentives to form a link with another green peripheral agent iff

$$\delta - \delta^3 < c \tag{37}$$

and this same condition ensures that a blue peripheral agent does not have incentives to form a link with another blue peripheral agent.

Following the same logic as in previous proofs, since each peripheral agent has one and only one bridge link, there is no exposure effect and none of these individuals has incentives to sever the link with the center in their community.

The green center does not have incentives to form a link with the blue center iff

$$\delta - \delta^3 - c < C \quad (38)$$

Because of symmetry, this same condition ensures that the blue center does not have incentives to form a link with the green center.

Observe that if (37) then (38) immediately follows.

The green center does not have incentives to form a link with a blue peripheral agent iff

$$\delta - \delta^2 - c < C$$

that holds by assumption. And this same condition ensures that the blue center does not have incentives to form a link with a green peripheral agent. Hence, because of mutual consent in link formation, we can ensure that no bridge link between the center of one community and a peripheral agent of the other is worth off.

A green peripheral agent does not have incentives to form a bridge link with another blue peripheral agent iff

$$\delta - \delta^3 - c - \frac{1}{4}C + \frac{1}{2}C < 0$$

which is equivalent to:

$$4 [c - (\delta - \delta^3)] > C \quad (39)$$

and, once more because of symmetry, this same condition ensures that a blue peripheral does not have incentives to form a bridge link with another green peripheral agent.

Hence, the required conditions are (36), (37) and (39).

(iv) Firstly, the two centers don't have incentives to sever the bridge link that connects them iff

$$\delta - \delta^3 - c > C \quad (40)$$

A blue peripheral individual does not have incentives to sever the link with a green peripheral one iff

$$\delta - \delta^3 - c - \frac{1}{n^G - 1} \frac{1}{n^B - 1} C + (n^G - 2) \left[\frac{1}{n^G - 2} \frac{1}{n^B - 1} - \frac{1}{n^G - 1} \frac{1}{n^B - 1} \right] C > 0$$

which is equivalent to

$$\delta - \delta^3 - c > 0$$

which trivially holds if (40) holds too. The same argument holds to show that a green peripheral individual does not have incentives to sever the link with a blue peripheral one.

A peripheral blue individual does not have incentives to form a link with the center of the other community iff

$$\begin{aligned} \delta - \delta^2 - c - \frac{1}{n^G} \frac{n^G - 1}{n^G} C - (n^G - 1) \left[\frac{1}{n^G} \frac{1}{n^B - 1} - \frac{1}{n^G - 1} \frac{1}{n^B - 1} \right] C &< 0 \\ \Leftrightarrow \delta - \delta^2 - c - \frac{1}{n^G} \frac{n^G - 1}{n^G} C + \frac{1}{n^B - 1} \frac{1}{n^G} C &< 0 \\ \Leftrightarrow \delta - \delta^2 - c < \frac{1}{n^G} \left[\frac{n^G - 1}{n^G} - \frac{1}{n^B - 1} \right] C \end{aligned}$$

which is a condition that is trivially satisfied given the assumption that $c > \delta - \delta^2$ and that the right hand side of this last inequality is strictly positive. An equivalent argument is valid for the incentives of a peripheral green not willing to form a link with the blue center. Hence, because of mutual consent, we do not have to check for the condition of a center of one of the communities not willing to form a link with a peripheral agent of the other community.

A peripheral agent does not have incentives to build a link with another peripheral of his own community because the direct benefit of this connection would be

$$\delta - \delta^2 - c < 0$$

and it would imply higher costs for the connections with the other community.

Finally, a peripheral blue has no incentives to sever the link with the blue center if

Hence, only the first of the inequalities, $\delta - \delta^3 - c > C$, is required for that network to be stable.

■

Proof of Proposition 6

The total surplus for complete segregation is equal to:

$$[n^G (n^G - 1) + n^B (n^B - 1)] (\delta - c)$$

while the total surplus for complete integration is given by:

$$\begin{aligned} n^G \left[(n^G - 1) (\delta - c) + n^B \delta - \left(c + \frac{(n^G - 1) (n^B - 1)}{(n - 1)^2} C \right) (n^B - 1) \right] \\ + n^B \left[(n^B - 1) (\delta - c) + n^G \delta - \left(c + \frac{(n^G - 1) (n^B - 1)}{(n - 1)^2} C \right) (n^G - 1) \right] \end{aligned}$$

Segregation is better if and only if:

$$\delta \leq \left[c + \frac{(n^G - 1) (n^B - 1)}{(n - 1)^2} C \right] \left(1 - \frac{n}{2n^G n^B} \right) \quad (41)$$

It is easy to check that $1 > \frac{n}{2n^G n^B}$ and, hence, that the upper bound is strictly positive.

For C large enough, segregation dominates integration. What happens when C is smaller? Let's take the smallest value C can take, i.e. $\delta + (n^B - 1)\delta^2 - c$, and see if integration dominates segregation. The condition (41) is now given by:

$$\delta > \left[c + \frac{(n^G - 1)(n^B - 1)[\delta + (n^B - 1)\delta^2 - c]}{(n-1)^2} \right] \left(1 - \frac{n}{2n^G n^B} \right)$$

where C has been replaced by $\delta + (n^B - 1)\delta^2 - c$. This is equivalent to:

$$\begin{aligned} & \delta \left[\frac{2(n-1)^2 n^G n^B + (n^G - 1)(n^B - 1)(2n^G n^B - n)}{2n^G n^B - n} \right] \\ & > \left[(n-1)^2 - (n^G - 1)(n^B - 1) \right] c + (n^G - 1)(n^B - 1)(n^B - 1)\delta^2 \\ & \Leftrightarrow \delta \left[\frac{2n^G n^B}{2n^G n^B - n} - \frac{(n^G - 1)(n^B - 1)}{(n-1)^2} \right] - \frac{(n^G - 1)(n^B - 1)^2}{(n-1)^2} \delta^2 \\ & > \left[1 - \frac{(n^G - 1)(n^B - 1)}{(n-1)^2} \right] c \\ & \Leftrightarrow c < \delta \left[\frac{\frac{2n^G n^B}{2n^G n^B - n} - \frac{(n^G - 1)(n^B - 1)}{(n-1)^2}}{1 - \frac{(n^G - 1)(n^B - 1)}{(n-1)^2}} \right] - \left[\frac{\frac{(n^G - 1)(n^B - 1)^2}{(n-1)^2}}{1 - \frac{(n^G - 1)(n^B - 1)}{(n-1)^2}} \right] \delta^2 \end{aligned}$$

We are in the range $c < \delta - \delta^2$. It is easy to verify that:

$$\frac{\frac{2n^G n^B}{2n^G n^B - n} - \frac{(n^G - 1)(n^B - 1)}{(n-1)^2}}{1 - \frac{(n^G - 1)(n^B - 1)}{(n-1)^2}} > 1$$

So, with the help of some algebra, we find that a sufficient condition is that

$$\frac{\frac{(n^G - 1)(n^B - 1)^2}{(n-1)^2}}{1 - \frac{(n^G - 1)(n^B - 1)}{(n-1)^2}} \leq 1$$

which is equivalent to (15). ■

Proof of Proposition 7 (from Appendix B)

(i) Let us first show that *complete integration* between communities is always an equilibrium network. There is *no* gain in utility for a green person to sever a link with a blue person, who is necessarily connected to the rest of the blue community, if:

$$\begin{aligned} & \delta - \delta^2 - c - \left[k + \left(\frac{n^G - 1}{n - 2} \right) \left(\frac{n^B - 1}{n - 2} \right) \right] C \\ & + (n^B - 1) \left[\left(\frac{n^B - 1}{n - 2} \right) \left(\frac{n^G - 1}{n - 3} \right) - \left(\frac{n^B - 1}{n - 2} \right) \left(\frac{n^G - 1}{n - 2} \right) \right] C \geq 0 \end{aligned} \quad (42)$$

The first term $\delta - \delta^2$ are the benefits derived from externalities of having a direct connection instead of an indirect connection with this blue person. The second term, $-c - \left[k + \left(\frac{n^G - 1}{n - 2} \right) \left(\frac{n^B - 1}{n - 2} \right) \right] C$, is the cost of forming the link with this blue person. Observe that, before forming the link, the proportion of green friends among all green person's friends is $\frac{n^G - 1}{n - 2}$, while the proportion of blue friends among all blue person's friends is $\frac{n^B - 1}{n - 2}$. The third term,

$$(n^B - 1) \left[\left(\frac{n^B - 1}{n - 2} \right) \left(\frac{n^G - 1}{n - 3} \right) - \left(\frac{n^B - 1}{n - 2} \right) \left(\frac{n^G - 1}{n - 2} \right) \right] C$$

is the indirect benefit derived from the diminishing costs of maintaining a link with a blue person, once this new link is formed. Before forming the new link, the proportion of green friends among all green person's friends is $\frac{n^G - 1}{n - 3}$. Once the new link is created, this proportion diminishes to $\frac{n^G - 1}{n - 2}$, and this implies a decrease in the cost of maintaining the link with the $n^B - 1$ blue persons from $c + \left(\frac{n^B - 1}{n - 2} \right) \left(\frac{n^G - 1}{n - 3} \right) C$ to $c + \left(\frac{n^B - 1}{n - 2} \right) \left(\frac{n^G - 1}{n - 2} \right) C$. The third term in (42) accounts for this difference in costs.

The inequality (42) is equivalent to

$$\delta - \delta^2 - c \geq \left[\left(\frac{n^B - 1}{n - 2} \right) \left[\frac{n^B - 1 - (n - 3)(n^G - 1)}{(n - 3)(n - 2)} \right] + k \right] C$$

The RHS of this inequality is negative iff:

$$k < \left[\frac{(n^B - 1)(n^G - 1)}{(n - 2)^2} - \frac{(n^B - 1)^2}{(n - 3)(n - 2)^2} \right]$$

A sufficient condition is thus:

$$k < \frac{(n^B - 1)(n^G - 1)}{(n - 2)^2}$$

Similarly, because of symmetry, the condition that guarantees that there is *no* gain in utility for a blue person to sever a link with a green person is

$$\delta - \delta^2 - c \geq \left[\left(\frac{n^B - 1}{n - 2} \right) \left[\frac{n^B - 1 - (n - 3)(n^G - 1)}{(n - 3)(n - 2)} \right] + k \right] C$$

which is always true if (45) is satisfied. As a result, *complete integration* between communities is always an equilibrium network if (45) is satisfied.

(ii) Let us show that *complete segregation* between communities is an equilibrium network. There is *no* gain in utility for a green person to establish a link with a blue person, who is necessarily connected to the rest of the blue community, if:

$$C > \frac{\delta + (n^B - 1) \delta^2 - c}{1 + k}$$

Similarly, there is *no* gain in utility for a blue person to connect to a green individual, who is necessarily connected to the rest of the green community, if:

$$\delta + (n^G - 1) \delta^2 - c < (1 + k) C$$

Since $n^G \geq n^B$, and because mutual consent is necessary, then condition (6) guarantees that there is complete segregation.

(iii) Let us find the condition that guarantees that there are no equilibrium for which each community is not fully connected. For that, we take the worst case scenario. The smallest benefit a blue person can obtain by making a link to another blue is $\delta - \delta^2$. The highest cost for a blue i to have a link with another blue is found by

$$\min_b \left\{ -c + n^G \left[\frac{b}{n^G + b} \times 1 - \frac{b+1}{n^G + b + 1} \times 1 \right] C \right\}$$

where $b \in [0, n^B - 2]$ is the number of blue friends of blue i . Observe that

$$\frac{b}{n^G + b} - \frac{b+1}{n^G + b + 1} = -\frac{n^G}{(n^G + b + 1)(n^G + b)} < 0$$

So

$$\min_b \left[\frac{b}{n^G + b} - \frac{b+1}{n^G + b + 1} \right] \Leftrightarrow b = 0$$

This implies that the worst case scenario is

$$\begin{aligned} \delta - \delta^2 - c - \frac{n^G}{(n^G + 1)} C &> 0 \\ \Leftrightarrow C &< \frac{(n^G + 1) (\delta - \delta^2 - c)}{n^G} \end{aligned}$$

If this is true then any blue will create a link with another blue. Doing the same procedure for greens, we obtain

$$C < \frac{(n^B + 1) [\delta - \delta^2 - c]}{n^B}$$

Since

$$\frac{(n^G + 1) [\delta - \delta^2 - c]}{n^G} < \frac{(n^B + 1) [\delta - \delta^2 - c]}{n^B}$$

Then the condition for both blues and greens is

$$C < \frac{(n^G + 1) (\delta - \delta^2 - c)}{n^G}$$

Putting together (6) and this condition leads to

$$\frac{\delta + (n^B - 1) \delta^2 - c}{1 + k} < C < \frac{(n^G + 1) (\delta - \delta^2 - c)}{n^G}$$

■

APPENDIX B: Oppositional identities when intercommunity costs are always larger.

In the previous section, we found that bi-partite networks were pairwise stable because there were no costs of becoming “green” for a blue person. In the equilibrium network described in Figure 8, the blue B_m “becomes” a green for other greens since the cost of interacting with her is just c . This is due to our assumption on the cost function which stipulates that the *intercommunity* cost is equal to the *intracommunity* cost as soon as one of the persons involved in the relationship has no friends of the same type. In the present section, we relax this assumption and assume instead the following *intercommunity* cost function for $\tau(i) \neq \tau(j)$:

$$c_{ij} = c + \left(k + e_i^{\tau(i)} e_j^{\tau(j)} \right) C \quad (43)$$

where $0 < k < 1$ (we still assume that $c_{ij} = c$ if $\tau(i) = \tau(j)$). With this new intercommunity cost function, a blue person can never become totally “green” for other greens because even if she has no blue friends, i.e. $e_i^{\tau(i)} = 0$, the cost of interacting with greens is $c + kC$, which is strictly greater than c , the cost for a green of interacting with other greens.

Proposition 7 *Assume $c < \delta - \delta^2$ and (43). If*

$$C < \frac{(n^G + 1) (\delta - \delta^2 - c)}{n^G} \quad (44)$$

holds, then any equilibrium network is such that each community is fully connected. In particular, a bipartite network (such that the one described in Figures 8 or 9) can never be an equilibrium. Furthermore, if

$$k < \frac{(n^B - 1) (n^G - 1)}{(n - 2)^2} \quad (45)$$

and

$$C > \frac{\delta + (n^B - 1) \delta^2 - c}{1 + k} \quad (46)$$

*hold, then both the network for which the blue and green communities are **totally integrated** and the one for which the blue and green communities are **completely segregated** are equilibrium networks.*

When the intercommunity cost function is given by (43), then each community forms a complete network if C is not too large. In that case, no bipartite network can emerge. This is because nobody can now become “like” someone from the other type and, therefore, the attractiveness of having only friends from the other community is much lower. Interestingly, when k is not too large and C high enough, each individual can either have links with all individuals (including those from the other community) or only links with her own community. Indeed, once the network is totally integrated, then nobody wants to delete a link because the gain is too low compared to the costs (this is because k is low enough). When the network is completely segregated, then because C is high enough, no individual wants to form a link with someone from the other community.