Group representation concerns and network formation

Melguizo Lopez, Isabel

Center for the Research and Teaching in Economics, Mexico City, Mexico

8 February 2022

Group Representation Concerns and Network Formation *

Isabel Melguizo[1]

[1]Economics Division, Centro de Investigación y Docencia Económicas (CIDE)
Correspondence: Isabel Melguizo, Economics Division, CIDE, Ciudad de México, Carretera México-Toluca 3655, Lomas de Santa Fe, 01210. Email: isabel.melguizo@cide.edu

Abstract

This paper studies processes of integration and segregation using a connections model in which individuals form valuable links that also entail a cost. Individuals belong to two different groups and care about whether their own group represents a sufficient fraction in their neighborhood. Concerns for representation promote the segregation of societies as even for small linking costs individuals do not link to different others because of the threat that their group become under-represented. For certain cost ranges, concerns for representation also determine efficient networks because forming links with members of the opposite group entails a utility loss due to under-representation.

Keywords: integration, segregation, representation concerns, homophily, welfare, pairwise stability
JEL classification: D6, D85, Z130

1 Introduction

Schelling (1969, 1971) postulates the striking result that segregation between groups arises even when individuals are as happy in a segregated as in a mixed society, as long as their own group is sufficiently represented in their neighborhood. Schelling's model brilliantly shows how mild individual preferences for representation can generate unintended consequences at the aggregate level. In this approach individuals care about the group to which others living in their local neighborhood, which is defined in terms of spatial proximity, belong. Thus, for any individual, a neighbor is someone located within a pre-specified distance. Then, in particular, individuals

*I am grateful to Mauricio Fernández, Antonio Jiménez-Martínez, Nikolas Tsakas, Yves Zenou and the seminar participants at Colmex, University of Granada, University of Málaga and ITAM for fruitful conversations. I also thank Luis Felipe Rosales for outstanding research assistance.

do not choose whom to be friends with. Of course, the relevance of spatial proximity is easily justified because individuals develop their activities in a physical space. However, it is also natural to think that individuals have some control over their social environment.

The main aim of this paper is to understand processes of integration and segregation when individuals, that belong to two different groups and exhibit preferences for how their own group is represented in their neighborhood, decide whom to be friends with.[1] In particular we study, to which extent *concerns for representation* promote segregated societies and how these concerns interplay with linking costs.

The idea that concerns for representation may influence linking decisions has been documented by Ingram and Morris (2007), who find how guests at mixers were more likely to join a group when it contained at least one same-race person. Concerns for representation signify that individuals care whether, at the collective level, their group is sufficiently visible. These concerns may entail that individuals do not want to establish relations with different others, which can be interpreted as a form of intolerance. Although most people in Western Europe express positive views of religious minorities, still discomfort with multiculturalism and intolerance are evident. In particular 43 % of the Italians said they would not accept Muslims as members of their families. The percentage was 36 % in the case of the United Kingdom.[2] As Aguiar and Parravano (2015) point out, intolerance has been always a major source of segregation and conflict. In this sense this paper investigates the relation between tolerance and the level of segregation in societies.

This paper relies on the well-known *symmetric connections model*, henceforth the *SC* model, by Jackson and Wolinsky (1996). While in the *SC* model individuals are ex-ante homogeneous in the current proposal individuals are ex-ante heterogeneous, in particular they belong to one of the two different groups that exist in the society. Concerns for representation materialize into that individuals' utility drops when they change their status from *over-representation*, that is, when their own group is present (weakly) above a *desired fraction* among their direct friends, to *under-representation*, when their own group is not sufficiently present among their direct friends, that is, below a desired fraction.[3] In a different framework, Alesina and La Ferrara (2000) study the relation between participation in social activities and the composition of groups. They propose a model in which individuals have some degree of intolerance to others. In particular, the individuals' utility decreases with the proportion of different others in the group they participate.

The main innovation of our paper is thus the introduction of a sort of meeting

[1]It is widely accepted that individuals choose their connections. In fact, the study of network formation has become essential for the understanding of relevant socio-economic outcomes. See Granovetter (1983), Bramoullé and Kranton (2007), and Calvó-Armengol et al. (2009) to cite a few relevant studies that incorporate network analysis.

[2]See the Western European Survey of the Pew Research Center at https://www.pewforum.org/2018/05/29/nationalism-immigration-and-minorities/.

[3]Other pieces of research based on Schelling's proposal model utility exhibiting this drop. See Fagiolo et al. (2007), Pancs and Vriend (2007), Fagiolo et al. (2009), Zhang (2011), and Grauwin et al. (2012). Also, regarding the segregation patterns between Blacks and Whites in the U.S., Card et al. (2008) conclude how the utility of whites seems to exhibit, in fact, a sharp drop beyond a certain fraction of blacks.

bias towards same-group others, due to the fact that individuals' linking decisions deliver not only link-based benefits but also benefits that have a collective meaning. More specifically, individuals are thinking how the representation status of their group is affected by their linking decisions, namely, whether they directly link to others of the opposite group and hence form *crossed links*, or to others of the same group and hence form *same-group links*. In a quite different setting, Currarini et al. (2009) introduce a condition that shares a somehow analogous spirit that the above mentioned preferences for representation, namely, higher returns of additional friendships when individuals enjoy a higher proportion of same-group acquaintances.[4]

Regarding indirect links, we follow Coleman (1988) and consider that for individuals, friends of friends are valuable because of their role in the transmission of resources, i.e., information, favors. In a nutshell, thus, individuals appreciate that their friends are popular.[5] For any individual we refer to the friends of their friends as indirect or distance-two connections/links. The value of these links is smaller than the one of direct links, the usual decay assumption. We consider that indirect links are equally valuable regardless of the group the individuals involved belong to.

Regarding the cost of establishing links, only direct links are (exogenously) costly. In particular, links between individuals of the two different groups, that is, crossed links, are the most expensive. That possibly reflects barriers to interracial contact as in Battu et al. (2007) and De Martí and Zenou (2017).[6] Links between individuals of the same group or same-group links, are then the cheapest ones.

The equilibrium concept is *pairwise stability*, introduced by Jackson and Wolinsky (1996). A network is pairwise stable, or simply stable, when no pair of individuals involved in a relation wants to terminate it and, for a pair of individuals not involved in a relation, at least one of the individuals does not want to establish it.

The results focus on the emergence of stable networks and on the relation between these stable networks and those that yield the highest sum of individual utilities, the efficient ones.

Regarding stability, the general intuition is that concerns for representation deter the emergence of crossed links and promote the formation of same-group links. Thus, stable structures tend to be more segregated than those that emerge in the absence of these concerns. The interplay between linking costs and the desired fraction in shaping stable networks is as follows:

First, when the cost of same-group links is sufficiently small, in equilibrium all individuals within each group are directly linked to each other. A group in which all its members are directly linked to each other is said to be a *connected group*.

[4] With this *same-group bias* the authors rationalize the fact that larger groups form a higher number of friendships, but they do not relate this condition to the emergence of more segregated societies.

[5] Currarini et al. (2017) also emphasize the role of distance-two connections and micro-found them. In their model distance-two friends congest the access to information, hence entailing negative externalities. In our case distance-two friends promote positive externalities.

[6] There is brief discussion in section 4 of why we opt for having linking costs that depend on the group individuals belong to.

When each group is connected, a variety of stable structures arises depending on the abundance of crossed links. In particular, and highlighting the two polar cases, for a sufficiently small cost of crossed links, the number of such links crucially depends on the desired fraction, i.e., a higher desired fraction leads to lesser connections with the opposite group and vice versa. When the cost of crossed links is sufficiently high, stable structures without crossed links emerge. As we will see, if concerns for representation were absent, the desired fraction would play no role and, for a sufficiently small cost of crossed links, everyone will directly link to everyone else. Some of the mentioned stable networks are consistent with the categorization of acculturation strategies due to Berry (1997). The author divides the acculturation strategies of immigrants in four dimensions: *integration*, *separation*, *marginalization*, and *assimilation*. Individuals that pursue a strategy of integration maintain interactions with their own group and also with the opposite one. Individuals that pursue a strategy of separation maintain interactions with their own group and avoid interactions with the opposite group.[7] In our approach, stable networks in which each group is connected and that have a certain abundance of crossed links may be interpreted as those in which individuals pursue a strategy of integration whereas stable networks with connected groups and without crossed links may be interpreted as those in which individuals pursue a strategy of separation.

Second, when the cost of same-group links starts to increase, stable networks in which groups are not connected emerge. Star-like networks, in which a central individual is connected to all the peripheral individuals and there are no other links, are a prominent example. Goeree et al. (2009) suggest how social relations are often based on the activity of few active central organizers in a friendship network. Also, an interesting finding is that even for sufficiently high same-group linking costs, individuals may want to be linked to same-group others even when these latter individuals do not bring any indirect benefit. While in the *SC* model that individuals provide each other with indirect benefits is necessary for these individuals to directly link when costs are sufficiently high, in our case it is precisely the threat of under-representation that the absence of same-group links entails, and the consequent utility loss, the reason as to why such links become appealing. In this latter respect, I contend that the model can also be applied to other settings as those of political coalitions. In particular, mainstream parties may partner with extremist parties, that provide the former ones with the opportunity to held a majority in elections, but apart from that there are no other benefits, as the ideal policies of the two types of parties probably differ. See Twist (2019).

Our proposal naturally relates to the phenomenon of *homophily*, the robust tendency of individuals to associate with similar others.[8] Thus, we also investigate the homophilous behavior of groups in stable networks. In particular, we analyze the proportion of same-group friendships within each group of individuals and how

[7]For completeness, a strategy of marginalization consists in that individuals look for isolation, and a strategy assimilation consists in that individuals maintain interactions mainly with different others.

[8]See, for instance, Shrum et al. (1988) for racial homophily and Ruef et al. (2003), for gender homophily. For a comprehesive survey on homophily see McPherson et al. (2001).

it depends on the primitives of the model.[9] Briefly, for stable networks in which both groups are connected we find that groups become more homophilous, namely, the proportion of same-group friendships increases, as the cost of crossed links increases. When the cost of a crossed link is sufficiently small, this proportion also increases with the desired fraction, that is, as individuals become intolerant to the presence of different others. Also, for certain parameter values, the proportion of same-group friendships within a group overcomes the relative representation of this group in society.

Regarding efficiency, we put some of our results into perspective with respect to the results delivered by the *SC* model. Among other elements of comparison, when the costs of forming crossed and same-group links are sufficiently small, in our case the network in which everyone is directly linked to everyone else, namely, the *integrated network*, may not be efficient when it is stable. Taking into account that in our case individuals are ex-ante heterogeneous, this network may be understood as the analogous counterpart of the complete network in the *SC* model, which is efficient when stable. The reason for this contrast precisely comes from the utility loss due to under-representation if individuals form many crossed links. It is important to mention, however, that when the integrated network is stable but not efficient, there is always a network which is both stable and efficient. This network is such that for no individual her group is under-represented in her neighborhood, and is called a *semi-integrated network of type* 1. More broadly, we also document the existing conflict between stability and efficiency for different cost levels.

This paper is related to the literature on network formation that analyzes versions of models of connections. Within this literature, there are two approaches: (i) those that focus on the emergence of pairwise stable networks, thus following Jackson and Wolinsky (1996), and (ii) those that focus on the emergence of Nash networks, as Galeotti et al. (2006) and Dev (2014), and that follow Bala and Goyal (2000). This paper is closer to the first approach.

Within the first approach, the *SC* model parsimoniously gives visibility to the concept of pairwise stability and to the tension between stability and efficiency. In it, direct links are equally valuable and costly, and indirect connections of any length are valuable. Thus, both our assumptions and our main focus are different. Johnson and Gilles (2003) consider exogenous costs that depend on the physical proximity between individuals located on a line. The notion of spatial proximity is absent in the current model, rather, decay is a measure of network connectedness. Jackson and Rogers (2005) propose a truncated version of a model of connections with exogenously heterogeneous individuals and heterogeneous linking costs, as in our case. However, their main aim, in contrast to ours, is to establish an economic-based reasoning for the emergence (and properties) of small-world networks. In De Martí and Zenou (2017) individuals are also ex-ante heterogeneous but the linking costs are endogenous and depend on the exposure to the opposite group. Their results on the relation between stability and efficiency are different from

[9] For this purpose we make use of the *inbreeding homophily index* developed by Coleman (1958).

ours. In particular, they focus only on the comparison between the network in which everyone is directly linked to everyone else and the network in which each group is connected and there are no crossed links.

The remainder of the paper is as follows. Section 2 sets the model. Section 3 presents the results. First, there is the analysis of stable networks and the homophilous behavior of groups. Second, the relation between stability and efficiency is established. Section 4 concludes. Section 5 contains the proofs.

2 The model

Individuals and groups. There is a finite population of $N = \{1, ..., n\}$ individuals, $n > 2$. Each individual belongs to one of the two groups, A and B. Let n_t be the number of individuals of group $t \in \{A, B\}$, with $n = \sum_t n_t$. The group an individual i belongs to is denoted $t(i) \in \{A, B\}$. We consider the case in which $n_A - 1 > n_B \geq 2$.

Network of relations. A network g is a set of links between the individuals in N. A direct link between individuals i and j is denoted ij. Then $ij \in g$ if and only if there exists a direct link between individuals i and j. The focus is on undirected networks, then the requirement is that $ij = ji$.

A crossed link is a link between individuals of different groups while a same-group link is a link between individuals of the same group.

The set of direct friends of individual i in g, also termed as the neighborhood of individual i in g, is $N_i(g) = \{j \in N | ij \in g\}$. Let $n_i(g)$ be the cardinality of this set. The set of direct friends of individual i in g that are of her same group is $N_i^s(g) = \{j \in N | ij \in g\ ,\ t(i) = t(j)\}$. Let $n_i^s(g)$ be its cardinality. The proportion of individuals of the same group than i in g, included i, in her neighborhood is $p_i^s(g) \equiv [n_i^s(g) + 1]/[n_i(g) + 1]$.[10]

There is a path from individual i to individual j in network g whenever one can go from i to j through the links of the network. The length of a path is the number of links involved in it.[11] The shortest path between i and j is the path that involves the lowest number of links. We define the distance between i and j as the length of the shortest path that connects them. The distance between i and j is denoted $d(i, j)$.[12]

The set of distance-two (or indirect) friends of individual i in network g is $I_i(g) = \{j \in N | d(i, j) = 2\}$.

The value of connections. Each individual derives a value of $\delta \in (0, 1)$ from each of her direct friends and a value of δ^2 from each of her indirect friends, no matter the group these individuals belong to.

The cost of connections. Direct links are (exogenously) costly. Formally, let $\overline{c}$ and $\underline{c}$ be two positive constants such that $\overline{c} > \underline{c}$. For each pair of individuals i

[10]When there is no ambiguity we omit the argument g in the mentioned proportion and simply write p_i^s.

[11]Formally, a path of length k between i and j in g is a sequence of individuals, $\{i_0, ..., i_k\}$, where $i_0 = i$ and $i_k = j$ and for $0 \leq p \leq k - 1$ we have that $i_p \neq i_{p+1}$ and $i_p i_{p+1} \in g$.

[12]When there is not path connecting two individuals we simply set $d(i, j) = \infty$.

and j, let $c_{t(i)t(j)}$ be the pair-dependent cost of a link. The assumption is that $c_{t(i)t(j)} = \underline{c}$ if $t(i) = t(j)$ and $c_{t(i)t(j)} = \overline{c}$ if $t(i) \neq t(j)$.

The utility function. For each individual i and given a network g, when the proportion of individuals of her same group in her neighborhood is below the desired fraction, that is, $p_i^s(g) < f \in (0, 1]$, that individual faces a loss of $d > 0$ as she is under-represented. Specifically, the utility of each individual i in network g is:

$$u_i(g) = \sum_{j \in N_i(g)} \delta + \sum_{k \in I_i(g)} \delta^2 - \sum_{j \in N_i(g)} c_{t(i)t(j)} - \mathbf{1}_{p_i^s(g)<f} d,$$

where $\mathbf{1}_{p_i^s(g)<f}$ takes value one when i is under-represented, and zero otherwise (when she is over-represented).

Two aspects are worth mentioning: first, the utility function is separable in the number of links an individual has. A considerable amount of the research in network formation considers separability.[13] A plausible alternative is to specify a convex cost function. That means that individuals find more costly an additional link the higher the number of links they already have. That definitely modifies the conditions on the costs for different stable structures to arise, as individuals with different number of links evaluate differently additional ones. The particular effect of the desired fraction, through the associated drop in utility, will however remain analogous to the case we consider.[14] Our specific cost structure allows to identify in a neat way the effect of the desired fraction. Second, what determines the value of connections is the shortest path between individuals. That means that if two individuals are directly linked, what counts is their direct link, regardless of whether they are also indirectly linked. In other words, individuals choose to link along the path with the highest reliability, the shortest one. Since communication decays with distance, that the shortest path is the most reliable seems to be a natural approach.[15]

The relation between population fractions and the desired fraction. Let $f_t = n_t/n$ be the fraction of individuals of group t in the population. Let f_{t-a} be the fraction of individuals of group t in the population when one individual of group A is

[13] See, Golub and Livne (2010) for separability in the value of links, De Martí and Zenou (2017) for separability in the value and cost of links, Iijima and Kamada (2017) for separability in the value of links and also in the cost, when the cost function is linear, and Baumann (2021) for separability in the value of non-costly links.

[14] For instance, and advancing the analysis, for a link between individuals i and j of different groups not to be formed, the requirement is that for at least one individual, say i, the marginal cost of the link is higher than its marginal benefit. The only difference with a convex cost function is that the marginal cost of a link is not $\overline{c}$, as in the current model, but $\Delta\overline{c}(n_i(g)) = \overline{c}(n_i(g) + 1) - \overline{c}(n_i(g))$ where $n_i(g)$ is the number of links i already has in g. This marginal cost clearly may vary by individual. That definitely affects the emergence of stable structures. Apart from that, the workings of the analysis are analogous. In particular, the drop in utility by $d > 0$ due to under-representation, only adds requirements on the value of $\Delta\overline{c}(n_i(g))$ for the link not to be formed, as in the current model. In particular, the loss d would imply that for a wider range of costs, the link will not be formed, as in the current case.

[15] Another interesting approach is the one taken by Bloch and Dutta (2009). In this case individuals choose to connect through the path with the highest reliability, not necessarily the shortest. In their model, link strength is an individual choice while in ours links are dichotomous. In general, we ignore then the possibility that the reliability of a path depends on other factors, as the strength of the links or the group the individuals in the path belong to. This latter approach is very interesting and remains for further research. For an interesting discussion on the multiple path specification, see Safi (2018).

deleted. The focus is on the case in which: (i) $f_A > f_{A-a} \geq f > f_{B-a} > f_B$ and (ii) $n_B/(n_B+1) \geq f$. Regarding (i), $f_A > f > f_B$ naturally accommodates to a situation in which the desired fraction may be the focal value of one half. Thus one group is a minority in the population and the other the majority. This choice does not affect the core conclusions on how the linking costs and the desired fraction shape the results. The inequalities $f_{A-a} \geq f > f_{B-a}$ and assumption (ii) are also set for clarity and do not qualitatively change the results.[16]

Equilibrium networks. The equilibrium concept is pairwise stability. To define it, let $g+ij$ denote the network g when the link ij is added. Analogously, let $g-ij$ denote the network g without the link ij. Pairwise stability is defined as follows:

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

In a pairwise stable network no individual gains from severing an existing link and no pair of individuals that are not directly linked, both gain when they link. Thus link formation is mutual consent while link deletion is unilateral.

We finally introduce useful definitions. Recall that a pair of individuals is directly linked in g if and only if $ij \in g$. Hence, directly linked individuals are those who have a direct relation. First of all, we introduce the notion of *connected group*, which is useful to further describe some networks of interest.

Definition 2. *Group $t = A, B$ is said to be connected if all the individuals that belong to that group are directly linked to each other.*

The following definitions categorize network structures and highlight prominent architectures.

Definition 3. *An integrated network is such that each group is connected and also, all individuals of different groups are directly linked to each other (thus, is everyone is directly linked to everyone else).*

Definition 4. *A segregated network is such that no pair of individuals of different groups is directly linked.*

A particular segregated network of interest is defined below:

Definition 4.1. *A CS network is a segregated network in which each group is connected.*

Definition 5. *A semi-integrated network is any structure not defined above.*

There are two semi-integrated networks of interest:

[16] In particular $f_{A-a} \geq f > f_{B-a}$ implies that in a network in which the individuals of each group are all linked among themselves, no individual of group A changes her representation status via link formation or deletion. Also, in a network in which everyone is linked, no individual of group B becomes over-represented by breaking one crossed link. Assumption (ii) means that an individual of group B, linked to all others of her same-group, does not become under-represented by linking to one individual of group A.

Definition 5.1. *A semi-integrated network of type* 1*,* SI_1*, is such that:* (i) *each group is connected and* (ii) *each individual of group* B *directly links to the number of individuals of group* A *to match the desired fraction* f*, in the sense that with less crossed links each individual of group* B *is over-represented whereas with more crossed links, she is under-represented.*

Definition 5.2. *A semi-integrated network of type* 2*,* SI_2*, is such that:* (i) *each group is connected and* (ii) *some individuals of group* B *match the desired fraction* f*, as in Definition* 5.1*, and, among the remaining individuals of group* B*, each of them directly links to all the individuals of group* A*.*

Figures 1 and 2 illustrate some of these definitions. In particular, figure 1 presents an integrated network (right) and a CS network (left). In the integrated network, the utility of each individual of group A is $5\delta - 3\underline{c} - 2\overline{c}$, whereas the utility of each individual of group B is $5\delta - \underline{c} - 4\overline{c} - d$. As we consider the case in which $n_A/n = 2/3 > f > n_B/n = 1/3$ and in this network for each individual i of group t, $p_i^s = n_t/n$, individuals of group B are under-represented, thus each suffers a loss of $d > 0$. Individuals of group A are, on the contrary, over-represented. In the segregated network no one is under-represented, the utility of each individual of group $t = A, B$ is $(n_t - 1)(\delta - \underline{c})$.

Figure 2 presents a SI_1 (left) and a SI_2 (right) network. Consider that $f = 1/2$. In the SI_1 network, for each individual i, $p_i^s \geq 1/2 = f$, thus everyone is over-represented in her neighborhood. Notice that each individual of group B directly links to two individuals of group A, to match $f = 1/2$. There are two individuals of group A that do not have crossed links but each accesses indirectly to the two individuals of group B. The utility of each of these individuals of group A is $3(\delta - \underline{c}) + 2\delta^2$. The utility of each of the other two, who have crossed links, is $5\delta - 3\underline{c} - 2\overline{c}$. The utility of each individual of group B is $3\delta + 2\delta^2 - \underline{c} - 2\overline{c}$, as, in particular, they access indirectly to two individuals of group A. In the SI_2 network, there is an individual i of group B who is over-represented, since she is directly linked to exactly two individuals of group A, hence matching the desired fraction, $f = 1/2$. Her utility is $3\delta + 2\delta^2 - \underline{c} - 2\overline{c}$. There is another individual j of group B who is under-represented, since she is directly linked to all the individuals of group A, and thus, $p_j^s = 1/3 < 1/2 = f$. Her utility is $5\delta - \underline{c} - 4\overline{c} - d$. Finally, all the individuals of group A are over-represented. Two of them are directly linked to the two individuals of group B and hence enjoy the same utility than in the integrated network. Each of the other two is directly linked to one individual of group B and enjoy an indirect link with the other. The utility of each of them is $4\delta + \delta^2 - 3\underline{c} - \overline{c}$.

3 Results

The results are divided between those referring to stable networks and those referring to the relation between stable networks and the networks that yield the highest sum of individual utilities. These networks are said to be efficient.

3.1 Equilibrium analysis

The following observation becomes important for the stability analysis. It allows us to divide stable networks depending on whether both groups are connected or not.

Lemma 1. *In a stable network both groups are connected if and only if* $\underline{c} \leq \delta - \delta^2$.

For such a small cost, establishing a same-group link is always profitable. If otherwise, this cost increases to become $\underline{c} > \delta - \delta^2$, a stable network cannot be such that both groups are connected, as there is always some pair who has incentives to break the link. As an example, in a CS network, in which both groups are connected and there are no crossed links, any pair of same-group individuals is willing to break since they become indirect friends anyway.

We first focus on stable networks in which both groups are connected. The general pattern is that in these stable structures crossed links progressively disappear as the associated cost increases and when the desired fraction increases, that is, when tolerance to the presence of different others decreases.

Proposition 1. *Let*

$$\underline{c} \leq \delta - \delta^2 \tag{1}$$

holds, then: .

1. *The integrated network is stable if and only if the cost of a crossed link is small enough, that is,* $\overline{c} \leq \delta - \delta^2$*. If* $\overline{c} \in (\delta - \delta^2 - d, \delta - \delta^2]$*, the* SI_1 *and* SI_2 *networks are stable in addition to the integrated network whereas if* $\overline{c} \leq \delta - \delta^2 - d$*, the integrated network is uniquely stable.*

The following networks in which both groups are connected, are uniquely stable:

2. *A network with* $z \in \{1, ..., n_B - 1\}$ *crossed links, each between a different individual of group* A *and a different individual of group* B*, if and only if the cost of a crossed link is intermediate, that is,* $\delta + (n_B - z - 1)\delta^2 < \overline{c} \leq \delta + (n_B - z)\delta^2$.
3. *Provided that* $f \leq n_B/[n_B + 2]$*, a network with* $z = n_B$ *crossed links, each between a different individual of group* A *and a different individual of group* B*, if and only if* $\delta - \delta^2 < \overline{c} \leq \delta$.
4. *The* CS *network if and only if the cost of a crossed link is high enough, that is,* $\overline{c} > \delta + (n_B - 1)\delta^2$.

SI_1 and SI_2 stable networks (point 1) reflect situations in which, apart from maintaining same-group relations, individuals of both groups relate to each other at different extents, depending on the desired fraction. As mentioned in the introduction, the pattern observed in SI_1 and SI_2 networks can be understood in terms of the strategy of integration defined by Berry (1997). According to this strategy, individuals maintain their cultural integrity through interactions with their own group, and are also part of the larger social network by interacting with the opposite group. In SI_1 networks each individual of group B directly links to the number of individuals of group A with which she matches the desired fraction. Relations with the opposite group are more abundant in SI_2 networks, since some individuals of group B relate to all the individuals of group A. Regarding the patterns of socialization of Mexican adolescents in the United States, Phinney et al. (2001) conclude how their behavior is consistent with a strategy of integration. It is important to mention that if concerns for representation were absent, and the cost of crossed links was also small enough, as in point 1 of the above proposition, the SI_1 and SI_2 networks would never arise as stable, as individuals will always want to directly link to everyone else. Thus, in this sense, these concerns promote the emergence of more segregated societies.

In the networks of point 2, there are few crossed links. In particular when $z = 1$, there are two individuals belonging to different groups who connect the two communities and hence enjoy a particular position: they connect otherwise disconnected groups. These individuals are covering a structural hole.[17]

The CS network (point 4) can be understood as the outcome of the strategy of separation defined by Berry (1997), by which individuals interact exclusively with members of their own group. An example of this socialization pattern is the Amish community in the U.S. The members of this community have strong family and social ties among themselves and desire to be isolated from the rest of the society. See Hostetler (1993). Also, Kromhout and Vedder (1996) conclude how the Antilleans in The Netherlands follow a behavior consistent with a strategy of separation due to the intolerance of the Dutch government.

To place the results within the literature, notice that the networks in which both groups are connected and that contain crossed links, are not stable in addition to the CS network. That is in contrast to De Martí and Zenou (2017) in which, in particular, the CS network and the integrated network are simultaneously stable for a wide range of parameter values. Notice also that when the integrated network is stable, is not uniquely stable for a specific cost range.[18] In particular, there are values of the cost of a crossed link such that the semi-integrated structures SI_1 and SI_2 are stable in addition to the integrated network. In contrast, in the SC model when the complete network, which can be understood as the analogous counterpart to our integrated network, is stable, is also uniquely stable. The potential multiplicity of stable structures is precisely due to the concern individuals have regarding their representation status. For instance, in the SI_1 networks the individuals of group B have incentives to directly link to the number of individuals of group A to match the desired fraction given the threat of under-representation. That opens a wedge on the cost, as described in point 1 of Proposition 1, for these networks to be stable in addition to the integrated network. Finally, notice that in the SI_1 and

[17]See Burt (1992) and Goyal and Vega-Redondo (2007) for the notion of structural hole and for the importance of links connecting communities for social capital.

[18]And outside this range it is uniquely stable.

SI_2 networks, as the desired fraction decreases, which can be interpreted as that individuals become more tolerant to the presence of different others, the number of crossed links increases.

The model is related to the notion of homophily, the pervasive phenomenon that similarity breeds connection. To relate more precisely stable networks with the homophilous behavior of groups, we use the inbreeding homophily index due to Coleman (1958), which measures the average number of same-group friendships a network has. In particular, for group $t = A, B$, the index is defined as:

$$IH_t = \frac{H_t - f_t}{1 - f_t},$$

where $H_t = s_t/[s_t + d_t]$ is an index of *relative homophily*. Here, s_t is the average number of friendships that individuals of the same group have. Analogously, d_t is the average number of friendships that individuals of a group have with others outside their group. Thus, H_t is the proportion of same-group friendships that are formed within a group.

Given f, SI_1 and SI_2 networks differ in the number of individuals of group B that are directly linked to all the individuals of group A. In the former networks this number is zero, whereas in the latter networks, is positive. We thus can easily parametrize these networks by defining $\mu \in \{1, ..., n_B\}$ as the number of the individuals of group B that directly link to the number of individuals of group A, to match the desired fraction. We denote the number of individuals A with which individuals of group B match the desired fraction, by f^m.[19] Thus, for $\mu = n_B$, a SI_1 network arises, otherwise, a SI_2 network arises. Finally, recall that in Proposition 1, z is the number of crossed links stable networks display. The results are as follows:

Lemma 2. *Let (1) hold. Then:*

1. *For $\delta - \delta^2 - d < \bar{c} \leq \delta - \delta^2$, the inbreeding homophily index is increasing in μ and f. Moreover:*
   - *(a) For $\mu = n_B$, the inbreeding homophily index is: (i) positive for group A and (ii) positive for group B if and only if $f^m < n_A(n_B - 1)/n_B$.*
   - *(b) For $\mu < n_B$, the inbreeding homophily index is: (i) positive for group A if and only if $\mu > n_B/(n_A - f^m)$ and (ii) positive for group B if and only if $\mu > n_A/(n_A - f^m)$.*
2. *For $\bar{c} > \delta - \delta^2$ the inbreeding homophily index is positive for both groups, increases as z decreases, and, is at least as high as in case 1 but smaller than one.*
3. *For $\bar{c} > \delta + (n_B - 1)\delta^2$, the inbreeding homophily index is one for each group.*

[19] Its precise value is $\lfloor n_B(1 - f)/f \rfloor$ and is (weakly) decreasing in f. Also, we omit the case $\mu = 0$, that is, the integrated network. In this case the inbreeding homophily index is slightly negative for small samples and tends to zero as sample size increases.

Some comments are in order. First, it is intuitive that both groups exhibit positive inbreeding homophily in SI_1 networks, case 1.(a), when the desired fraction is high enough. In this case no individual of group B wants to directly link to many individuals of group A. In SI_2 networks, case 1.(b), the index is positive for both groups when the number of individuals of group B who do not directly link to all the individuals of group A, is sufficiently high. As the desired fraction increases, the individuals of group B do not want to directly link to many individuals of group A, thus, the number of individuals of group B who directly link to others in group A to match the desired fraction, can be small. The fact that there are individuals of group B who establish few crossed links compensates the presence of those other individuals of group B who directly link to all the individuals of group A. Second, the networks with z crossed links, case 2, have the minimum value of the index when the highest number of crossed links have been formed, that is, $z = n_B$. Notice that in terms of the index, a network with $z = n_B$ crossed links is equivalent to a SI_1 network in which each individual of group B has one crossed link. Thus, in the former network the index is at least as high than in any SI_1 network. Third, stable networks display positive inbreeding homophily for both groups, under certain parameters values. That is consistent with the findings in Currarini et al. (2009). Notice also that in any stable network the larger group A forms more total friendships per capita than the smaller group B, which is also in line with the findings in Currarini et al. (2009). Moreover, the smaller group B forms more different-group friendships per capita than the larger group A. This is in line with Blau (1977) who argues that, due to the reciprocal nature of ties, there will be more cross-group friendships per capita for the smaller group than for the larger group. Moreover, by comparing cases 1 with 2 and 2 with 3, observe that the inbreeding homophily index increases with the cost of a crossed link. Finally, in the SI_1 and SI_2 networks the index increases with the desired fraction.

As the cost of a same-group link increases to become $\underline{c} > \delta - \delta^2$, the two groups cannot be connected, see Lemma 1. The following result provides some characteristics that stable networks display in this case.[20] To describe the results, I find useful to say that an individual i receives *exclusive indirect connections* from another individual j when i can only access to these indirect connections via a link with j. In other words, if i and j are not directly linked, i does not enjoy these indirect connections through a direct link with any other individual $k \neq j$.

Proposition 2. *Let*

$$\underline{c} \in (\delta - \delta^2, \delta] \tag{2}$$

hold. Then, in a stable network:

1. *A pair of individuals of the same group is directly linked if and only if one of the following conditions holds: (i) they do not access each other indirectly. (ii) They access each other indirectly and each individual: either receives exclusive*

[20] The analysis of crossed links considers that $\delta^2 > d$. The case in which $\delta^2 < d$ implies that only $\overline{c} \in (\delta - \delta^2, \delta]$ is relevant. The results are analogous to the ones exposed and omitted for concreteness.

indirect connections from the other individual or, if she does not, she becomes under-represented by breaking the link.

Let $\delta^2 > d$. Then:

2. *For $\overline{c} \in (\delta - \delta^2, \delta - d]$, a pair of individuals of different groups is directly linked if and only if one of the following conditions holds: (i) they do not access each other indirectly. (ii) They access each other indirectly and each individual receives exclusive indirect connections from the other individual.*

3. *For $\overline{c} \in (\delta - d, \delta]$, a pair of individuals of different groups is directly linked if and only if one of the following conditions holds: (i) they do not to access each other indirectly and either: none of them becomes over-represented by breaking the link or, some of them becomes over-represented by breaking the link, but she loses exclusive indirect connections she was receiving from the other individual. (ii) They access each other indirectly, each individual receives exclusive indirect connections from the other individual and either: none of them becomes over-represented by breaking the link or, some of them becomes over-represented by breaking the link, but she loses more than one exclusive indirect connection she was receiving from the other individual.*

4. *For $\overline{c} > \delta$, if a pair of individuals of different groups is directly linked, then each of them receives exclusive indirect connections from the other individual.*

The statements of the above proposition exhaustively cover the cases that may arise regarding link formation. The general message is however direct: concerns for representation introduce a channel by which in stable networks it would be easier to observe more same-group links and less crossed links than if these concerns were absent. More specifically, the number of situations under which same-group links arise amplifies. In particular, in case 1, the last situation in which an individual does not lose indirect connections by breaking the link but faces the threat of under-representation, would be absent without concerns for representation. Analogously, concerns for representation restrict the number of situations under which crossed links arise. Without these concerns, case 2 reduces to the condition that $\overline{c} \in (\delta - \delta^2, \delta)$ and case 3 vanishes. Thus, for a crossed link to exist, there are no requirements on the minimum number of exclusive indirect connections individuals must receive.

Case 1 also implies that same-group individuals must be at distance of, at most, two. Notice that they may be linked despite of accessing each other via an indirect link. That happens not only because they may provide each other with indirect connections that otherwise, if they break, no individual is able to enjoy, but also because there may be a threat of becoming under-represented by breaking the link. Regarding links between individuals of different groups, incentives to link vary with the cost. If the cost is sufficiently small, as in case 2, individuals must also be at a distance of, at most, two. Notice that the upper bound on the cost takes into account the possibility that by breaking the link individuals become over-represented. As the cost increases, additional requirements for a crossed link to exist are progressively added: for instance, in case 3, if someone gets the premium of becoming over-represented by breaking a crossed link, then she must be losing indirect connections for such a link to exist in a stable network. Importantly, case 3 implies that individuals of different groups may prefer not to stay linked even if they are at a distance higher than two. That happens when the crossed link implies a transition to the under-representation status and at least one of the individuals involved does not get indirect connections due to this link. This case is illustrated below in the star-like networks. Case 4 establishes the necessary condition that linked individuals of different groups must be receiving indirect connections from each other.

Star-like structures are prominent examples. Anecdotal evidence suggests how some social relations are usually based on the activity of few active central organizers who are central in a friendship network.[21] The following lemma describes stable star-like networks.

Lemma 3. *Under (2), the following networks are stable:*

1. *For $\delta^2 > d$, a star where the central individual belongs to group A (figure 3) if and only if:*

   (i) $\overline{c} \leq \delta$, *for* $f \in [0, 1/2] \cup (2/3, 1]$,

   or

   (ii) $\overline{c} \leq \delta$ *and* $\underline{c} > \delta - \delta^2 + d$, *for* $f \in (1/2, 2/3]$.

2. *Two stars, each encompassing all individuals of the same group, with one crossed link between the central individuals of each star (figure 4, right), if and only if* $\overline{c} \in (\delta, \delta + (n_B - 1)\delta^2]$ *for* $f \in [0, 2/3]$ *or* $\overline{c} \in (\delta - d, \delta + (n_B - 1)\delta^2]$, *otherwise.*

3. *Two stars, each encompassing all all individuals of the same group (figure 4, left), if and only if* $\overline{c} > \delta + (n_B - 1)\delta^2$.

The desired fraction plays an important role in these networks. It is useful to recall that the formation same-group links favors that individuals are over-represented, and its break is detrimental for this purpose. The contrary happens with crossed links. Consider case 1 and focus first on point (*ii*), where the desired fraction is intermediate. In this case, the individuals of group B pass from under to over-representation when they link. Thus, same-group link formation is very appealing. As a consequence, the associated per link cost must be sufficiently large, otherwise, we should observe them being linked, and the star would be not stable. Absent the possibility of over-representation, the condition will just be, $\underline{c} > \delta - \delta^2$, which is always satisfied, thus these individuals would never link.

[21] As Goeree et al. (2009) point out co-author networks display structures in which there are few well-connected researches collaborating with many others.

In case 1.(i), the desired fraction is either sufficiently low so that no individual changes her representation status via same-group linking or sufficiently high so that the individuals of group B do not change their under-representation status when they link.[22]

In case 2, the lower bound on the cost of a crossed link changes with the desired fraction. That is due to the potential change of the representation status. Notice that the peripheral individuals of the two groups do not access each other indirectly. By linking, any individual k of group B would become under-represented whenever $f > 2/3$. Additionally, no individual l of group A, with which k would link, provides k with exclusive indirect connections. That is so because, k already has access to the central individual of group A via the link with the central individual of group B. Notice then that there is no reason to sustain a such a link. In particular, none of the requirements in cases 3 or 4 of the above Proposition 2, for these individuals to stay linked, is satisfied. Observe also that when the desired fraction is high enough the range on the cost of a crossed link for which this network is stable, is wider than if the desired fraction is low. An interpretation is that more tolerant individuals, those facing a small desired fraction, would have formed crossed links for cost ranges in which more intolerant individuals, those facing a high desired fraction, would not have done it. Thus, high desired fractions induce stable networks that are less integrated. Case 3 presents a stable segregated network, the conditions are reminiscent to those for the stability of the CS network.

Finally, when $\underline{c} > \delta$, the stability results partially resemble the ones in Jackson and Wolinsky (1996).[23] In their case, in non-empty stable networks linked individuals must have at least two links. That is true in our case when stable networks have no crossed links. However, when stable networks have crossed links, that does not have to be true. As being linked to same-group individuals favors that the over-representation status is attached, individuals may find profitable to maintain same-group links, even when they do not provide indirect benefits. The result is as follows:

Proposition 3. *Let $\underline{c} > \delta$. Then, in any stable network that has crossed links, all linked individuals involved in a crossed link must have at least two links. Moreover, there are values of the desired fraction for which there exist stable networks with crossed links in which some individuals have just one link with another individual of their same group.*

Figure 5 illustrates a network with one crossed link. Among the individuals of group B, the central one becomes under-represented if she breaks a same-group link. That opens a wedge for the cost of a same-group link to be such that it is profitable to maintain such links with individuals that do not bring indirect benefits.

This latter observation may have different applications. In particular, it may help to understand and rationalize the formation of political coalitions, in which

[22] The group of the central individual is, in general, inessential for this characterization.

[23] See their Proposition 2.

mainstream parties may partner with extremist parties. Extremist parties may provide mainstream parties with the opportunity to held a majority in elections, but apart from that there are no other benefits as the ideal policies of the two types of parties probably differ.[24]

3.2 The relation between stability and efficiency

To discuss the conflict between stability and efficiency, we consider, as it is common in the literature, that the value of a network is the sum of individual utilities. A network is efficient when its value is not smaller than the value of any other network.

In general, the conflict materializes into that efficient networks may not be stable and in that stable networks are not efficient. We analyze the relation between stability and efficiency in our context, and put it in perspective with regard to the SC model.

Recall that f^m is the number of crossed links any individual of group B who is linked to all same-group others, needs to have to match the desired fraction. Let:

$$\overline{c}(e; \delta, d) \equiv \delta - \delta^2 - \frac{d}{2e}, \tag{3}$$

where $e \in \{1, ..., n_A - f^m\}$ is the number of extra crossed links any individual of group B, who is linked to all other same-group individuals, may have beyond f^m. Expression (3) defines the cost of a crossed link such that the addition of e extra crossed links, does not affect the value of the network. The result is as follows:

Proposition 4. *Let $\overline{c} \leq \delta - \delta^2$, then:*

1. *For each desired fraction f, all the SI_1 networks yield the same value. These networks are uniquely efficient whenever $\overline{c} > \overline{c}(n_A - f^m; \delta, d)$.*
2. *For each desired fraction, the integrated network is uniquely efficient when there exists e^*, such that $\overline{c} < \overline{c}(e^*; \delta, d)$.*

Given that the cost of a same-group link is sufficiently low, in efficient networks each group is connected. As the cost of a crossed link is also small enough, in efficient networks individuals of group B must be linked to, at least, f^m individuals of group A. The question is whether adding extra crossed links increases network value. The answer may be clear if one notices that the drop in utility an individual of group B suffers when she becomes under-represented, is the only channel by which network value may decrease. As this drop is fixed, when the per link cost of a crossed link is sufficiently high, in particular, such that it overcomes the value of creating the maximum number of crossed links, the best thing to do to maximize network value is not to have any extra crossed link. That is considered in case 1 in the proposition above, thus SI_1 networks are efficient. If, on the contrary, there is a number of crossed links such that, for this number the value of the network

[24] See Twist (2019) for the support of this thesis and for empirical evidence on the electoral gains of far right parties in Western Europe in the last four decades.

increases with respect to the value of the SI_1 network, the value of the network is in fact maximized when all individuals of group B have the highest number of crossed links as possible. Then, the integrated network is uniquely efficient. That is represented in case 2. Overall, notice that in contrast to the SC model, the integrated network may not be efficient when it is stable. It is important however to emphasize here that for the given cost structure, there is always a stable network which is efficient, either a SI_1 network, or the integrated network.

In connecting this efficiency result with the existing literature, Buechel and Hellmann (2012) establish that in a network formation model with positive externalities in which utility is convex (in own current links), pairwise stable networks that are not efficient cannot be *over-connected*. Briefly, convexity establishes that the marginal utility of a group of existing links is higher than the sum of the marginal utilities that each of these link provides.[25] An over-connected network is one whose value can increase when removing some existing links. In case 1 of Proposition 4, SI_2 networks are not efficient but over-connected, since by removing links from them, the efficient SI_1 network results. The reason is that in our context, utility is not convex in current links. More precisely, for the bound on the cost of a crossed link in point 1 of Proposition 4, any individual of group B linked to all individuals of group A in a SI_2 network, is interested in breaking a group of crossed links to become over-represented, while she is not interested in breaking each of these links, one at a time. In other words, the marginal utility of this group of crossed links is negative, while the marginal utility of each of these links is positive.

In contrast, in the SC model utility is convex in current links. Then, networks that are not efficient cannot be over-connected.[26] That is also the case in a slightly different version of the model in De Martí and Zenou (2017).

The following result illustrates the conflict between stability and efficiency when the cost of a same-group link is still small enough but the cost of a crossed link is higher than in the proposition above.

Proposition 5. *Let* $\underline{c} \leq \delta - \delta^2$*, then:*

1. *When a network in which both groups are connected and there are* $z \in \{1, ..., n_B - 1\}$ *crossed links, each between a different individual of group* A *and a different individual of group* B*, is uniquely stable, is not efficient.*

2. *There is a range for the cost of a crossed link such that a network in which both groups are connected and there are* $z = n_B$ *crossed links, each between a different individual of group* A *and a different individual of group* B*, is uniquely efficient but not stable.*

3. *There is a range for the cost of a crossed link such that the* CS *network is uniquely stable but not efficient.*

[25] See also Calvó-Armengol and İlkılıç (2009) for the analogous notion of sub-modularity and its implications.
[26] For an illustration, see example 1 in Buechel and Hellmann (2012).

Overall, this result establishes that there are cost ranges for which stability and efficiency cannot be reconciled. For instance, in case 1, the range of the cost of a crossed link such that the network with z crossed links is (uniquely) stable, happens to be inconsistent with the range such that this network is efficient.

In the *SC* model, the conflict between stability and efficiency relies on the existence of a cost range such that no stable network is efficient.[27] Under such a cost range the star is uniquely efficient but never stable. This type of conflict also arises here with the difference that: in case 1 efficient networks must be such that both groups are connected, and hence cannot be star-shaped.

The last result establishes that when the cost of a same-group link is high enough, there are star-shaped networks that are uniquely efficient but not stable.

Proposition 6. *Let $\underline{c} > \delta$, then there is a range of costs such that the network consisting on two stars, each encompassing all same-group individuals, is uniquely efficient but not stable.*

For a sufficiently high cost of same-group links, the network consisting on two stars each encompassing all same-group individuals, see figure 4, is not stable. In particular, the central individual of each star does not want to maintain same-group links. However, there are instances in which this network is uniquely efficient. That happens when the cost of crossed links is high enough so that such links do not increase network value.

4 Discussion

This paper analyzes how preferences for representation shape processes of integration and segregation. Group representation concerns play an important role in determining stable structures in which both groups are connected. When this is not the case, concerns for representation allow to obtain results that are qualitatively different from those in the *SC* model. See, for instance, Proposition 3. The role of group representation concerns and heterogeneous costs in the relation between stability an efficiency is emphasized. In Proposition 4, although there is always a stable network which is efficient, this network is not always the integrated one. Proposition 5 illustrates the impossibility of reconciling stability and efficiency.

There is one aspect which is important to mention: the model is qualitatively similar in many respects to one in which links are equally costly, that is, when forming a link costs $c > 0$ for each of the individuals involved, no matter the group they belong to. However, with equally costly links, the emergence of stable structures in which both groups are connected there are no crossed links, is precluded. Specifically, the results of Proposition 1 for $c \leq \delta - \delta^2$, collapse to that only the integrated network and the SI_1 and SI_2 structures are stable. That implies that parsimonious results on how increasing linking costs determines the emergence of stable structures in which crossed links progressively disappear as the

[27] That happens when $\delta < c < \delta + 2^{-1}(n-2)\delta^2$.

associated cost increases, cannot be obtained. Moreover, for high values of the desired fraction, the SI_1 and SI_2 networks exhibit very weak homophilous behavior. For $c \in (\delta - \delta^2, \delta]$, stable structures in which one individual of each group is linked to all similar others must also contain at least one crossed link, for any desired fraction. Thus, high segregation patterns are, in many instances, difficult to rationalize. This precludes, in particular, the rationalization of star-shaped networks without crossed links in figure 4.

The model with heterogeneous costs is able to parsimoniously deliver results in which we can connect linking costs to the emergence of stable structures in which crossed links progressively disappear. Thus, it naturally delivers strong segregation patterns. That is in line with the evidence supporting the persistence of situations in which groups are highly isolated. McPherson and Smith-Lovin (1986) find a high degree of sex segregation in voluntary organizations. In particular, around 68 % of the organizations studied were composed exclusive by males or females. As Neto et al. (2005) document, segregationist/separatist strategies carried out by individuals are quite common. Ananat (2011) argues how high levels of segregation have persisted in the U.S from the late sixties to the present, despite of the flourishment of civil rights movements. Moreover, there are groups, as the already mentioned Amish community in the U.S, whose members have strong ties among themselves, and whose desire is to be isolated for the rest of the society.

5 Proofs

Proof of Lemma 1. A same-group link brings, at least, $\delta - \delta^2 - \underline{c} \geq 0$. Thus a stable network must be such that each group is connected for $\underline{c} \leq \delta - \delta^2$.

If $\underline{c} > \delta - \delta^2$ in stable networks it cannot be that each group is connected. To see that, consider a network in which each group is connected. There are two cases: (i) if the network does not have crossed links, any individual i of group A will break with another individual j of the same group. As j remains part of $i's$ indirect connections in the case of a break, individual i gains $-\delta + \delta^2 + \underline{c} > 0$ by breaking with j. (ii) If the network contains crossed links is because some individual of group A, say k, is linked to some individuals of group B. Thus, k already has access to all the individuals of group B, directly or indirectly. As any other individual j of group A remains part of $k's$ indirect connections in the case of a break, individual k gains $-\delta + \delta^2 + \underline{c} > 0$ by breaking with j. ■

Proof of Proposition 1. Let $\underline{c} \leq \delta - \delta^2$ hold. By Lemma 1, in stable networks each group is connected. Thus, the focus is on crossed links.

The proof is divided in four lemmas that are combined to establish the results of Proposition 1. Lemma A sets the upper bound on $\bar{c}$ for which the CS network is not stable. Lemma B establishes the condition under which at least z crossed links, each between different individuals, form. Lemma C summarizes the linking decisions of individuals of group B, considering the threat and disutility of under-representation. Lemma D considers the linking decisions of individuals of group A. A combination of these lemmas gives raise to the results of Proposition 1.

Lemma A. For $\bar{c} \leq \delta + (n_B - 1)\delta^2$, the CS network is not stable.

Proof of Lemma A. Let $\bar{c} \leq \delta + \delta^2(n_B - 1)$. Departing from the CS network, for any individual of group A, the benefit of forming a crossed link, i.e., the value of one direct connection plus $n_B - 1$ indirect connections, overcomes the cost. That is also the case for any individual of group B, since $n_A > n_B$. Hence, this link forms and the CS network is not stable.

Let us use the expression *distinct pair* to refer to a pair conformed by two individuals, one belonging to group A and the other to group B, such that none of them is already involved in a crossed link.

Lemma B. Let $\bar{c} \leq \delta + (n_B - z)\delta^2$ with $z \in \{1, ..., n_B\}$. Then, at least z distinct pairs link in a stable network. In particular, for a fix z, going iteratively over $m \in \{1, .., z - 1\}$, the distinct pair m links and then, the distinct pair $m + 1$ also links.

Proof of Lemma B. For $\bar{c} \leq \delta + (n_B - z)\delta^2$ the first distinct pair links, since $\bar{c} \leq \delta + (n_B - 1)\delta^2$ holds as $z \geq 1$. When this pair is linked and $z \geq 2$, a second distinct pair will also link as $\bar{c} \leq \delta + (n_B - 2)\delta^2$. Specifically, the two individuals involved receive $\delta^2 > 0$ due to one indirect connection without their link, whereas with the link they get, at least, $\delta + (n_B - 1)\delta^2 - \bar{c} \geq \delta^2$.[28] In general, for $\bar{c} \leq \delta + (n_B - z)\delta^2$ and each considered $z \in \{1, ..., n_B\}$, let $m \in \{1, ..., z - 1\}$. Then, the pair $m = 1$ forms a link. When this pair forms a link, the distinct pair $m + 1 = 2$ also emerges, and so on iteratively. Thus, when the pair $m = z - 1$ forms a link, the pair $m + 1 = z$ also emerges. Hence, for $\bar{c} \leq \delta + (n_B - z)\delta^2$ at least z crossed links, each between a distinct pair, emerge in a stable network.

Lemma C. The following holds regarding the linking decisions of individuals of group B:

1. For $\bar{c} \leq \delta - \delta^2 - d$ each individual of group B is willing to link to all the individuals of group A, regardless of the threat of under-representation.

2. For $\delta - \delta^2 - d < \bar{c} \leq \delta - \delta^2$: (i) if an individual of group B is over-represented, she is willing to link to the number of individuals of group A to match the desired fraction f, in the sense of Definition 5.1, given the threat of under-representation. (ii) If an individual of group B is under-represented, she is willing to link to all the individuals of group A, as there is no threat of under-representation.

3. For $\bar{c} > \delta - \delta^2$ no individual of group B who already has a crossed link, is willing to link to another individual of group A, even if there is no threat of under-representation with this second link.

4. For $\delta - \delta^2 < \bar{c} \leq \delta + (n_B - z)\delta^2$, $z = \{1, ..., n_B - 1\}$, in a network in which z distinct pairs are linked, each individual of group B who does not have crossed

[28] The individual of group B receives more since $n_A > n_B$.

links, is willing to link to an individual of group A, as there is no threat of under-representation.

Proof of Lemma C. We prove points 1 to 4.

1. For $\overline{c} \leq \delta - \delta^2 - d$ each individual of group B will link to all individuals of group A because the cost is small enough. In particular the cost takes into account the disutility in case of becoming under-represented so that the value of each link overcomes it.

2. Let $\delta - \delta^2 - d < \overline{c} \leq \delta - \delta^2$ and $f^m \equiv \lfloor (1-f)n_B/f \rfloor$ be the number of crossed links any individual of group B, who is linked to all others of their own group, needs to have to match f, in the sense of Definition 5.1. Consider that some individual i of group B is linked to $\tilde{n}_A < f^m$ individuals of group A. By linking to another individual of group A, i gains since she remains over-represented and $\overline{c} \leq \delta - \delta^2$ holds. Consider that some individual j of group B is linked to $\tilde{n}_A \in (f^m, n_A)$ individuals of group A. By linking to another individual of group A j gains since she remains under-represented and $\overline{c} \leq \delta - \delta^2$ holds. Thus, any individual of group B will be willing to link all individuals of group A or to f^m individuals of group A. In fact, for the proposed bounds on $\overline{c}$, given the threat of under-representation and the associated loss in utility: (i) any individual of group B who is linked to less than f^m individuals of group A, and hence over-represented, has incentives to link to up to f^m of them. With an additional link beyond f^m, she becomes under-represented and hence loses, given the lower bound on the cost. (ii) Any individual of group B who is linked to more than f^m individuals of group A is already under-represented. Given the upper bound on the cost she has incentives to link to all the individuals of group A.

3. Let $\overline{c} > \delta - \delta^2$. Then, for $n_B/(n_B+2) < f$ and $\delta - \delta^2 - d < \overline{c}$ an individual of group B who has one crossed link and becomes under-represented by forming a second one incurs in a loss, given the lower bound on the cost. For $n_B/(n_B+2) \geq f$ and $\delta - \delta^2 < \overline{c}$ an individual of group B who has one crossed link and remains over-represented by forming a second one incurs in a loss, given the lower bound on the cost. Thus, even when there is no threat of under-representation, she is not interested in forming such a second link.

4.- Consider that $\delta - \delta^2 < \overline{c} \leq \delta + (n_B - z)\delta^2$ and that $z = \{1, ..., n_B - 1\}$ crossed links, each between different individuals, have been formed. Then, an individual of group B with no crossed links will be interested in forming one. With one crossed link the individual of group B does not become under-represented, since by assumption $n_B/[n_B+1] \leq f$. Thus, for her there is no threat of under-representation. As it holds that $\overline{c} \leq \delta + (n_B - z)\delta^2 < \delta + (n_A - 1 - z)\delta^2$, given that, by assumption, $n_A - 1 > n_B$, a crossed link benefits her.

Lemma D. The following holds regarding the linking decisions of individuals of group A.

1. For $\overline{c} \leq \delta - \delta^2$ each individual of group A is willing to link to all the individuals of group B.

2. For $\overline{c} > \delta + (n_B - z - 1)\delta^2$, $z \in \{1, ..., n_B\}$, in a network in which z distinct pairs are linked, no individual of group A is willing to link to two individuals of group B.

Proof of Lemma D. We prove points 1 and 2.

1.- Consider that $\overline{c} \leq \delta - \delta^2$. As $f_A > f$ no individual of group A becomes under-represented by linking to any number of individuals of group B. Thus, she is willing to link to all of them.

2.- Consider that, $\overline{c} > \delta + (n_B - z - 1)\delta^2$. Then, no individual of group A, with no crossed links, has incentives to form one. As this individual already has access to z indirect connections with the individuals of group B, she obtains $z\delta^2$. If she links to an individual of group B, she obtains $\delta + (n_B - 1)\delta^2 - \overline{c} < z\delta^2$. Thus, she is not interested in such a link. An individual of group A who has a crossed link, has even less incentives to form another, as $\delta - \delta^2 < \overline{c}$ holds. Thus, no individual of group A will link to two different individuals of group B.

We obtain the following results:

Proposition 1.1. By Lemma C and Lemma D, the integrated network is stable if and only if $\overline{c} \leq \delta - \delta^2$. To see that consider point 2 in Lemma C, in particular the case that each individual of group B is willing to link to all individuals of group A. By point 1 in Lemma D each individual of group A is also willing to link to all individuals of group B, thus all the crossed links form and hence the integrated network is stable. On the contrary, for $\overline{c} > \delta - \delta^2$, point 3 in Lemma C, no individual of group B who has a crossed link will establish a second one. As link formation is mutual consent, the integrated network cannot be stable. By point 1 in Lemma D and point 2 in Lemma C, the SI_1 and the SI_2 networks are stable for $\delta - \delta^2 - d < \overline{c} \leq \delta - \delta^2$. By point 1 in Lemma D each individual of group A is willing to link to all the individuals of group B. As link formation is mutual consent, the resulting networks depend on the behavior of the individuals of group B. Point 2 in lemma C in fact implies that the only networks that can be stable in addition to the integrated network, for the proposed cost range, are the SI_1 and the SI_2 networks. Individuals of group B are willing to link to f^m individuals of group A or to all of them. In the case in which each individual of group B is willing to link to f^m individuals of group A, a SI_1 arises as stable. In particular, it is stable for $\delta - \delta^2 - d < \overline{c} \leq \delta - \delta^2$. The reason is that the lower bound on $\overline{c}$ precludes crossed link formation as individuals of group B are not interested in establishing more than f^m crossed links given the disutility of under-representation. The upper bound precludes crossed link deletion by any individual, not matter her group. In the case in which each of the $\mu \in (0, n_B)$ individuals group B is willing to link to f^m individuals of group A and each of the remaining $\mu - n_B$ individuals is willing to link to all the individuals of group A, a SI_2 network arises as stable. Again, $\delta - \delta^2 - d < \overline{c} \leq \delta - \delta^2$, precludes both crossed link formation and deletion. Recall that under the upper bound on $\overline{c}$, the integrated network is also stable. Also, notice that the integrated network is uniquely stable if and only if $\overline{c} \leq \delta - \delta^2 - d$. In this case, by point 1 in Lemmas C and D everyone wants to link to everyone else, thus no other network is stable. The conclusion is therefore that if $\overline{c} \in (\delta - \delta^2 - d, \delta - \delta^2]$ the SI_1 and SI_2 networks are stable in addition to the integrated network whereas if $\overline{c} \leq \delta - \delta^2 - d$ the integrated network is uniquely stable.

Propositions 1.2 and 1.3. Lemma B establishes the condition $\overline{c} \leq \delta + (n_B - z)\delta^2$, such that a stable network must have at least z crossed links, each between a distinct pair. This condition is necessary and sufficient. If it holds the individuals of each group are interested in linking in the way proposed. Thus, z crossed links form by mutual consent. If it does not hold, that is, if $\overline{c} > \delta + (n_B - z)\delta^2$, there is some individual of group A who is not willing to form the z^{th} link. Without the link she obtains $(z-1)\delta^2$. By establishing the link she gets less, $\delta + (n_B - 1)\delta^2 - \overline{c} < (z-1)\delta^2$.

We have that: (i) Proposition 1.2. For $\overline{c} \leq \delta + (n_B - z)\delta^2$ and when $z \leq n_B - 1$ crossed links have been formed, an individual of group B with no crossed links will be, in fact, interested in forming another crossed link, see point 4 in Lemma C. As link formation is mutual consent, it turns out that exactly z crossed links form if $\delta + (n_B - z - 1)\delta^2 < \overline{c}$, which is the condition in Lemma D point 2, by which no individual of group A wants to form a crossed link. In the opposite case the crossed link will form. Thus the proposed lower bound on the cost, $\delta + (n_B - z - 1)\delta^2 < \overline{c}$, is also necessary and sufficient for no more than z crossed link to emerge. We therefore conclude that $\delta + (n_B - z - 1)\delta^2 < \overline{c} \leq \delta + (n_B - z)\delta^2$ is necessary and sufficient for a stable network to have exactly $z \leq n_B - 1$ crossed links. (ii) Proposition 1.3. For $\overline{c} \leq \delta$ and $z = n_B$, each individual of group B has one crossed link. In this case, the individuals of group B remain over-represented with the second crossed link, that is, $n_B/[n_B + 2] \geq f$. The necessary and sufficient condition for a network with exactly $z = n_B$ crossed links to be stable is, in an analogous way than in case (i) above, $\delta - \delta^2 < \overline{c} \leq \delta$.

Uniqueness of the networks of points 2 and 3 of Proposition 1. When a network with exactly z crossed links is stable, is also uniquely stable. Specifically, for $\delta + (n_B - z - 1)\delta^2 < \overline{c} \leq \delta + (n_B - z)\delta^2$ and $z \leq n_B - 1$ a network with $z' \neq z$ distinct pairs linked cannot be stable. If $z' > z$, by Lemma D the individual of group A will be better off by breaking a link. If $z' < z$, by Lemma B, at least z link must emerge. Moreover, no network in which some individuals have more than one crossed link is stable as there is always a gain for breaking one of them, provided that $\overline{c} > \delta - \delta^2$. For $\delta - \delta^2 < \overline{c} < \delta$ and $z = n_B$, a network with $z'' < z$ distinct pairs linked cannot be stable, by an analogous argument than above, based on Lemma B. Also as above, no network in which some individual has more than one crossed link is stable.

Proposition 1.4. Under the complementary condition to the one in Lemma A, that is, $\overline{c} > \delta + \delta^2(n_B - 1)$, the CS network is stable as no crossed link can be formed. Therefore, the CS network is stable if and only if $\overline{c} > \delta + \delta^2(n_B - 1)$. In this case is also uniquely stable.

■

Proof of Lemma 2. Let $\underline{c} \leq \delta - \delta^2$. Case 1. For $\delta - \delta^2 - d < \bar{c} \leq \delta - \delta^2$, stable networks are of type SI_1 and SI_2. Let $\mu \in \{1, ..., n_B\}$ be the number of individuals of group B who are linked f^m individuals of group A. Thus, when $\mu = n_B$ we are in a SI_1 network, otherwise, we are in a SI_2 network. In the latter network, each of the $n_B - \mu$ individuals of group B, is linked to all individuals of group A. The relative homophily index is thus $H_t = n_t(n_t - 1)/[n_t(n_t - 1) + \mu f^m + (n_B - \mu)n_A], t = A, B$. Recall that $f^m \leq n_A - 2$. Thus, H_t and IH_t, $t = A, B$, increase with μ. Notice also that f^m is decreasing in f. Thus, both indexes increase with f. For $\mu = n_B$, $H_A - f_A = n_B n_A[n_A - 1 - f^m]/n[n_A(n_A - 1) + n_B\tilde{f}] > 0$, since $f^m \leq n_A - 2$. Also, $H_B - f_B = [n_A(n_B - 1) - n_B f^m]/n[(n_B - 1) + \tilde{f}]$. That expression is positive if and only if $f^m < n_A(n_B - 1)/n_B$. Direct algebra shows that, for $\mu < n_B$, $H_A - f_A > 0$ if and only if $\mu > n_B/(n_A - f^m)$ and $H_B - f_B > 0$ if and only if $\mu > n_A/(n_A - f^m)$.

Case 2. For $\bar{c} > \delta - \delta^2$, stable networks have $z \in \{1, ..., n_B\}$ crossed links and the relative homophily index is $H_t = n_t(n_t - 1)/[n_t(n_t - 1) + z]$, $t = A, B$. In the case in which $z = n_B$, it turns out that $H_t - f_t > 0$, $t = A, B$ if and only if $n_A(n_B - 1) > n_B$, which always holds. Thus the inbreeding homophily index is positive for both groups. Notice that the inbreeding homophily index for a network with $z = n_B$ crossed links is the same than the index of a network of case 1 in which $f^m = 1$ and $\mu = n_B$. Thus, in the case that $z = n_B$, the index is at least as high than in the stable networks of case 1. It is direct to see that as z decreases, the relative index, and hence, the inbreeding homophily index increase. Finally, as $z \geq 1$, $H_t < 1$ and hence, $IH_t < 1$.

Case 3. For $\bar{c} > \delta + (n_B - 1)\delta^2$, the CS network is uniquely stable. It has no crossed links, thus the indexes take a value of one for each group. ■

Proof of Proposition 2. I find useful to say that an individual i receives *exclusive indirect connections* from another individual j when i can only access to these indirect connections via a link with j. I use x to denote the minimum number of these connections individuals i and j enjoy when they are linked to each other.

To analyze when a link is formed, we consider the marginal gain associated to that particular link. The rest of the utility individuals gain in a given network due to other possible existing links is thus kept fixed and normalized to zero.

First, consider same-group link formation:

Point 1. Let $\delta - \delta^2 < \underline{c} \leq \delta$. In a stable network a pair of same-group individuals, i, j, is linked if: (i) they do not access each other via an indirect link. With the link each individual gains, at least, $\delta + x\delta^2 - \underline{c} > 0$, $x \geq 0$. Without the link they gain 0 in the best case in which they do not become under-represented. Thus, they are linked. (ii) They access each other via an indirect link and each individual: (a) either receives exclusive indirect connections from the other, or (b) if she does not, she becomes under-represented by breaking the link. In case (a), with the link the individual gains, at least, $\delta + x\delta^2 - \underline{c} > 0$, $x \geq 1$. Without the link, the gain is $\delta^2 > 0$ in the best case in which the individual does not become under-represented. Thus, the individual prefers to stay linked. In case (b), with the link the gain is

$\delta - \underline{c} > 0$. Without the link, the gain is $\delta^2 - d$, due to under-representation. Thus, as long as $\delta - \delta^2 + d > \underline{c}$, the individual prefers to stay linked.

If none of these conditions is satisfied, i and j access each other via an indirect link, but some of them, say i, does not receive exclusive indirect connections from j, neither becomes under-represented by breaking the link. As $\underline{c} > \delta - \delta^2$, i is better off without the link. Thus, a same-group link emerges if and only if some of the above conditions holds.

Second, let $\delta^2 > d$ and consider crossed link formation:

Point 2. Let $\overline{c} \leq \delta - d$. Then, in a stable network a pair of individuals, k, l, of different groups, is linked if: (i) they do not access each other via an indirect link. With the link each individual gains, at least, $\delta + x\delta^2 - \overline{c} > 0$, $x \geq 0$. Without the link they gain, at most, $d > 0$, if they become over-represented. As $\delta + x\delta^2 - \overline{c} > d$, they prefer to stay linked. (ii) They access each other via an indirect link and each of them receives exclusive indirect connections from the other. With the link each individual gains, at least, $\delta + x\delta^2 - \overline{c} > 0$, $x \geq 1$. Without the link each gains, in the best case of becoming over-represented, $\delta^2 + d > 0$. As $\delta + x\delta^2 - \overline{c} > \delta^2 + d$, they prefer to stay linked.

If none of these conditions is satisfied, k and l access each other via an indirect link but some of them, say k, does not receive exclusive indirect connections from l. As $\overline{c} > \delta - \delta^2$, then k is better off without the link. Thus, a crossed link emerges if and only if some of the above conditions holds.

Point 3. Let $\overline{c} \in (\delta - d, \delta]$. Then, in a stable network a pair of individuals, k, l, of different groups, is linked if: (i) they do not to access each other via an indirect link and none of them becomes over-represented by breaking the link. With the link they gain, at least, $\delta + x\delta^2 - \overline{c} > 0$, $x \geq 0$. Without the link they gain 0. Thus, they are linked. Also, they are linked if when some of them becomes over-represented by breaking the link, she loses exclusive indirect connections. With the link this individual gains, at least, $\delta + \delta^2 - \overline{c} > 0$. Without the link she becomes over-represented and enjoys $d > 0$. Given that $\delta^2 > d$, $\delta + \delta^2 - \overline{c} > d$, thus this individual stay linked. If none of these conditions holds, they do not access each other via an indirect link and some of them becomes over-represented by breaking the link without losing indirect connections. Then, she is better off without the link, as by breaking it she gains $-\delta + d + \overline{c} > 0$. Thus the link does not form.

(ii) They access each other via an indirect link, each individual receives exclusive indirect connections from the other individual and either: (a) none of them becomes over-represented by breaking the link or, (b) some of them becomes over-represented by breaking the link, but loses more than one exclusive indirect connection. In case (a) none of them becomes over-represented by breaking the link. With the link they gain, at least, $\delta + x\delta^2 - \overline{c} > 0$, $x \geq 1$. By breaking the link the gain is $\delta^2 > 0$. As, $\delta + x\delta^2 - \overline{c} > \delta^2$, they stay linked. In case (b), someone becomes over-represented by breaking the link but loses more than one exclusive indirect connection. Let this individual be k. With the link k gains, at least, $\delta + 2\delta^2 - \overline{c} > 0$. By breaking the link k becomes over-represented and gains $\delta^2 + d > 0$. Given that $\delta^2 > d$,

$\delta + \delta^2 - d - \overline{c} > 0$, thus k stay linked. If none of these conditions holds, they access each other via an indirect link, but some of them, say k, does not receive indirect connections from the other individual l. Then k is better off without the link, because she gains, at least, $-\delta + \delta^2 + \overline{c} > 0$ by breaking it, in the worst case in which she does not become over-represented. Thus the link does not form. The other possibility is that k becomes over-represented by breaking the link but loses just one exclusive indirect connection from l. Then k gains with the link $\delta + \delta^2 - \overline{c} > 0$ whereas if she breaks she gains, $\delta^2 + d > 0$. As $\overline{c} > \delta - d$ she is better off without the link. Hence, the link does not form.

Thus overall, if and only if one of the above conditions in cases (i) and (ii) holds, a crossed link emerges.

Point 4. For $\overline{c} > \delta$, any pair of individuals must necessarily receive from each other exclusive indirect connections. Otherwise the link is not profitable for them. By breaking it, they both gain even if they do not become over-represented.

■

Proof of Lemma 3. Point 1. Consider a star with a central individual of group A. She does not break any link if and only if $\overline{c} \leq \delta$ holds. That implies that no peripheral individual of group A breaks with her. No pair of peripheral individuals of group A links if and only if $\underline{c} > \delta - \delta^2$. No peripheral individual of group A wants to form a crossed link if and only if $\overline{c} > \delta - \delta^2$ holds under $f \leq 2/3$, since A remains over-represented. In contrast, $\overline{c} > \delta - d - \delta^2$ must hold under $f > 2/3$, since she becomes under-represented. For group B there are several cases:

Case 1. Let $f \leq 1/2$. Hence, when $\overline{c} \leq \delta + (n-2)\delta^2$, $\underline{c} > \delta - \delta^2$ and $\overline{c} > \delta - \delta^2$, an individual of group B does not break with the central individual neither wants to link to any other individual. Taking into account the conditions for the group A, the conclusion is that the star is stable if and only if $\underline{c} > \delta - \delta^2$ and $\overline{c} \in (\delta - \delta^2, \delta]$ hold.

Case 2. Let $1/2 < f \leq 2/3$. For no individual of group B to break with the central individual, $\overline{c} \leq \delta - d + (n-2)\delta^2$ must hold. Also, individuals of group B become over-represented by linking. To prevent these links, $\underline{c} > \delta + d - \delta^2$ must hold. Individuals of group B remain under-represented when linking to individuals of group A, thus $\overline{c} > \delta - \delta^2$ must hold to prevent these links. It already holds given the above conditions for group A. Thus, the star is stable if and only if $\underline{c} > \delta + d - \delta^2$ and $\overline{c} \leq min\{\delta, \delta - d + (n-2)\delta^2\} = \delta$, as $d < \delta^2$.

Case 3. Let $f > 2/3$. Recall that for the individuals of group A not to link among themselves, $\underline{c} > \delta - \delta^2$ must hold. Then, no individual of group B establishes a new link since she remains under-represented. Thus, the star is stable if and only if $\underline{c} > \delta - \delta^2$ and $\overline{c} < min\{\delta, \delta - d + (n-2)\delta^2\} = \delta$ hold, since $d < \delta^2$.

Point 2. Consider two stars, each encompassing all individuals of the same group, with a link between their central individuals. Regarding same-group links, the individuals of group A do not link or break among themselves if and only if $\underline{c} \in (\delta - \delta^2, \delta]$ holds. Under this condition, individuals of group B do not break or link among themselves. For the crossed links there are two cases:

Case 1. Let $f \leq 2/3$. The central individuals do not break with each other if and only if $\bar{c} \leq \delta + (n_B - 1)\delta^2$ holds. No pair of peripheral individuals of different group links if and only if $\bar{c} > \delta$ holds. That implies that the central individual of group A (respectively, group B) does not want to link to any peripheral individual of group B (respectively, of group A). Thus, the network is stable if and only if $\bar{c} \in (\delta, \delta + (n_B - 1)\delta^2]$ and the conditions on $\underline{c}$ hold.

Case 2. Let $f > 2/3$. For the central individuals their not breaking condition is the one in case 1. Links between peripheral individuals of different group cause them to become under-represented. Thus, $\bar{c} > \delta - d$ must hold to prevent these links. Therefore, the network is stable if and only if $\bar{c} \in (\delta - d, \delta + (n_B - 1)\delta^2]$ and the condition for $\underline{c}$ holds.[29]

Point 4. Consider two stars, each encompassing all individuals of the same group, without crossed links. Neither same-group link severance nor creation happens if and only if $\underline{c} \in (\delta - \delta^2, \delta]$ holds. The central individuals do not link if and only if $\bar{c} > \delta + (n_B - 1)\delta^2$ holds. That implies that no other crossed link forms. ∎

Proof of Proposition 3. As $\bar{c} > \underline{c} > \delta$, by point 4 in the proof of Proposition 2, any pair of individuals of different groups that are linked, must provide each other with exclusive second-order connections, otherwise they break. That implies that each linked individual must have at least to links. For the second part of the statement in the proposition, see figure 5. ∎

Proof of Proposition 4. For $\bar{c} \leq \delta - \delta^2$ in efficient networks both groups must be connected and such that each individual of group B links to, at least, f^m individuals of group A, who also gain with these links. Fix f and let e be the extra links to individuals of group A, beyond f^m, that any individual of group B may establish. For each individual of group B, these e links cause a change in the value of the network of:

$$- d + e(\delta - \delta^2 - \bar{c}) + e(\delta - \delta^2 - \bar{c}). \tag{4}$$

Suppose that at $e = n_A - f^m$, (4) < 0. Then at any $e' < e$, the change in value is even more negative. Thus, SI_1 networks, all of which yield the same value, are efficient.[30] That is so because the network value increases when each individual of group B has the least possible number of crossed links. If there exists a e^* such that (4) > 0 when evaluated at it, then the integrated network is uniquely efficient. That is so because the network value increases when each individual of group B has the highest number of crossed links as possible. ∎

Proof of Proposition 5. Let $\underline{c} \leq \delta - \delta^2$. Thus, in efficient networks both groups must be connected.

[29] Notice that as $\bar{c} > \underline{c} > \delta - \delta^2$, the central individual of group A does not link to any peripheral individual or to the central individual of group B, neither the central individual of group B links to any peripheral individual of group A, even when such an individual of group B does not change her representation status with this link.

[30] To see why all the SI_1 networks yield the same value, consider one of these SI_1 networks. By rearranging a number x crossed links from one individual of group A to another, the former loses $x(\delta^2 - \delta + \bar{c})$ and the latter gains $x(\delta - \delta^2 - \bar{c})$ and nothing else changes.

Point 1. Let g^z be a network with $z \in \{1, ..., n_B - 1\}$ crossed links, each involving a different individual of group A and a different individual of group B. Let $v(g^z)$ be its value, net of same-group links. Then:

$$v(g^z) = 2z(\delta - \overline{c}) + z(n_A - 1)\delta^2 + z(n_B - 1)\delta^2 + z(n_A - z)\delta^2 + z(n_B - z)\delta^2.$$

It holds that $v(g^z) > v(g^{z+1})$ whenever $\overline{c} > \delta + \delta^2(n - 2z - 2)$ holds. The bound decreases with z, thus $v(g^z) > v(g^{z+1}) \rightarrow v(g^{z'}) > v(g^{z'+1})$, $z' > z$. Analogously, $v(g^z) < v(g^{z+1})$ whenever $\overline{c} < \delta + \delta^2(n - 2z - 2)$ and $v(g^{z-1}) < v(g^z)$ whenever $\overline{c} < \delta + \delta^2(n - 2(z - 1) - 2)$. The bound increases as z decreases, thus $v(g^{z-1}) < v(g^z) \rightarrow v(g^{z''-1}) < v(g^{z''})$, $z'' < z$. Thus g^z yields the highest value, among the networks with crossed links, each involving different individuals, for:

$$\overline{c} \in (\delta + (n - 2z - 2)\delta^2, \delta + (n - 2(z - 1) - 2)\delta^2). \tag{5}$$

Notice the following aspects: (i) a network with $a \in \{1, ..., n_B\}$ crossed links yields the highest value when the maximum number of individuals has just one crossed link. To see that, depart from g^z, in which every individual has one crossed link, and diminish the number of individuals with one crossed link, keeping z fixed. The value of direct connections due to crossed links in g^z is $2z(\delta - \overline{c})$, regardless of who links to whom. Thus, focus on second-order connections. Start with individuals of group A. Diminishing the number of individuals of group A with one crossed link, is done by making one individual of group A unconnected to individuals of group B, and endow another individual of group A, with two crossed links. The unconnected individual of group A loses no second-order connections in the best case (when each individual of group B has a link to other individuals of group A). The individual of group A with the extra link loses δ^2. The individuals of group A not involved in this rearrangement remain the same. Thus, overall, the individuals of group A lose with this rearrangement. With respect to individuals of group B, those connected to some individuals of group A remain the same (the only thing that changes is the individual of group A they are linked to and that does not matter). As crossed links now concentrate on a smaller number of individuals of group A, if there are individuals of group B without crossed links, they lose second-order connections. Thus, overall, such rearrangement cannot increase network value. Consider now individuals of group B. The case in which some individual of group B becomes unconnected to an individual of group A, follows an analogous reasoning than for the unconnected individuals of group A. The individual of group B with one extra link may become under-represented, thus losing $d > 0$, apart from δ^2. Then, such rearrangements decrease network value. This reasoning can be repeated departing from any network with crossed links, by concentrating them in a smaller number of individuals. Hence, among the networks with a fixed number of crossed links a, the one in which the maximum number of individuals has just one crossed link, yields the highest value. Thus, under (5), g^z yields higher value than any other network with a number of crossed links $a \in \{1, ..., n_B\}$. Thus, in efficient networks, crossed links must be each between a different individual of group A and a different individual of group B. (ii) The lower bound in (5) implies that $\overline{c} > \delta$. Thus, when g^z yields the highest value among the networks with crossed links, each involving a different individuals, it also yields higher value than any other network with more than n_B crossed links. More precisely, by adding crossed links, the individuals involved lose, at least, $\delta - \delta^2 - \overline{c}$ and the individuals not involved are not affected. (iii) The upper bound in (5) implies than the network in which $z = 1$ yields higher value than the CS network. Then, if a network with $\tilde{z} > 1$ crossed links yields the highest value among the networks with $\hat{z} \neq \tilde{z}$ crossed links, it also yields higher value than the CS network. Thus, $(i) - (iii)$ imply that, under (5), g^z is uniquely efficient.

By Proposition 1.2, g^z is stable if and only if $\overline{c} \in (\delta+(n_B-z-1)\delta^2, \delta+(n_B-z)\delta^2]$ holds. Recall that it is also uniquely stable. That bound is incompatible with $\overline{c} \in [\delta + (n - 2z - 2)\delta^2, \delta + (n - 2(z - 1) - 2)\delta^2]$, under which g^z is efficient. In particular, notice that $n_A - z$ takes the minimum value of $n_A + 1 - n_B$ when $z = n_B - 1$. In particular, $n_A + 1 - n_B > 2$. Thus, $n_A - z > 2$ for any z. It implies that $n_B - z < n - 2z - 2$. Thus, when g^z is uniquely stable, is never efficient.

Point 2. Let g^z, $z = \{n_B - 1, n_B\}$, be defined as in point 1. It holds that $v(g^{n_B-1}) < v(g^{n_B})$ whenever $\overline{c} < \delta + \delta^2(n_A - n_B)$. This expression is analogous to the upper bound of (5) for $z = n_B$. As stated in point 1, this bound increases as z decreases, thus reducing crossed links results in lower value. Recall also that by 1.(i), for a fixed number of crossed links, the networks that yield the highest value are those in which these links are each between different individuals. Thus, under $\overline{c} < \delta + \delta^2(n_A - n_B)$ the network with n_B crossed links yields higher value than any other network with a lesser number of crossed links, regardless of who links to whom. When $\overline{c} > \delta - \delta^2$, g^{n_B} yields higher value than any other network with more than n_B crossed links. Summing up, for $\overline{c} \in (\delta - \delta^2, \delta + \delta^2(n_A - n_B))$, g^{n_B} is uniquely efficient. Moreover, when $\overline{c} \in (\delta, \delta+\delta^2(n_A-n_B))$ this network is uniquely efficient, but not stable, since stability requires that $\overline{c} \leq \delta$. See Proposition 1.

Point 3. By Proposition 1, the CS network is stable if and only if $\overline{c} > \delta + (n_B - 1)\delta^2$ holds. In this case it is also uniquely stable. When further $\overline{c} < \delta+(n-2)\delta^2$, this network is not efficient, since adding a crossed link increases network value. Thus, under $\overline{c} < \delta + (n - 2)\delta^2$ efficient networks necessarily contain crossed links. ■

Proof of Proposition 6. For $\overline{c} > \delta + (n - 2)\delta^2$ no network with crossed links is efficient. For $\underline{c} > \delta - \delta^2$, within the non-empty networks without crossed links, the one consisting on two stars, each encompassing all same-group individuals, brings the highest value. The proof is analogous to the one of Proposition 1.(ii) in Jackson and Wolinsky (1996), since the same logic applies to each component consisting on same-group individuals. The value of such a network is:

$$2(\delta - \underline{c})(n - 2) + [(n_A - 1)(n_A - 2) + (n_B - 1)(n_B - 2)]\delta^2.$$

When $\underline{c} < \delta + [2(n - 2)]^{-1}[(n_A - 1)(n_A - 2) + (n_B - 1)(n_B - 2)]\delta^2$ this network yields positive value and thus, it also yields more value than the empty network.

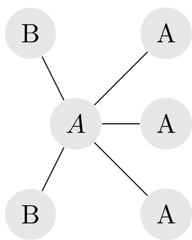
Recall that for the network consisting on two stars to be stable, $\underline{c} \leq \delta$ must hold. Thus, for $\underline{c} \in (\delta, \delta + [2(n-2)]^{-1}[(n_A - 1)(n_A - 2) + (n_B - 1)(n_B - 2)]\delta^2)$ and $\overline{c} > \delta + (n-2)\delta^2$, such network is uniquely efficient but not stable. ■

Figures

Figure 1: A CS network (left) and an integrated network (right)

Figure 2: A SI_1 (left) and a SI_2 network (right)

Figure 3: A star where the central individual belongs to group A



References

Aguiar, F. and A. Parravano (2015). Tolerating the intolerant: homophily, intolerance, and segregation in social balanced networks. *Journal of Conflict Resolution 59*(1), 29–50.

Figure 4: Two stars encompassing each all same-group individuals, without crossed links (left) and with one crossed link (right)

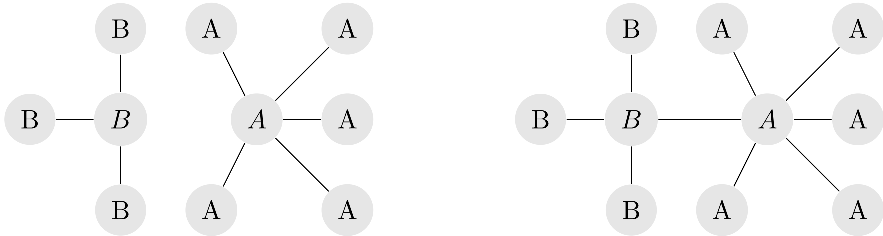
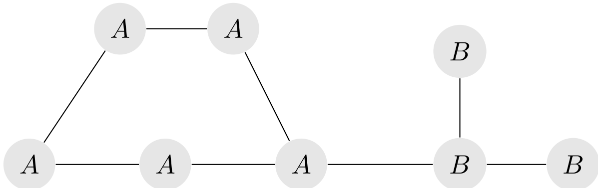


Figure 5: Let $f \in (2/3, 3/4)$. For $\underline{c} \in (\delta, min\{\delta + \delta^2, \delta + d\}]$ and $\bar{c} \in (\delta + 2\delta^2 - d, \delta + 2\delta^2]$ the following network is stable



Alesina, A. and E. La Ferrara (2000). Participation in heterogeneous communities. *The quarterly journal of economics 115*(3), 847–904.

Ananat, E. O. (2011). The wrong side (s) of the tracks: The causal effects of racial segregation on urban poverty and inequality. *American Economic Journal: Applied Economics 3*(2), 34–66.

Bala, V. and S. Goyal (2000). A noncooperative model of network formation. *Econometrica 68*(5), 1181–1229.

Battu, H., M. Mwale, and Y. Zenou (2007). Oppositional identities and the labor market. *Journal of Population Economics 20*(3), 643–667.

Baumann, L. (2021). A model of weighted network formation. *Theoretical Economics 16*(1), 1–23.

Berry, J. W. (1997). Immigration, acculturation, and adaptation. *Applied psychology 46*(1), 5–34.

Blau, P. M. (1977). *Inequality and heterogeneity: A primitive theory of social structure*, Volume 7. Free Press New York.

Bloch, F. and B. Dutta (2009). Communication networks with endogenous link strength. *Games and Economic Behavior 66*(1), 39–56.

Bramoullé, Y. and R. Kranton (2007). Public goods in networks. *Journal of Economic Theory 135*(1), 478–494.

Buechel, B. and T. Hellmann (2012). Under-connected and over-connected networks: the role of externalities in strategic network formation. *Review of Economic Design 16*(1), 71–87.

Burt, R. S. (1992). *Structural holes.* Harvard university press.

Calvó-Armengol, A. and R. İlkılıç (2009). Pairwise-stability and nash equilibria in network formation. *International Journal of Game Theory 38*(1), 51–79.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009). Peer effects and social networks in education. *The Review of Economic Studies 76*(4), 1239–1267.

Card, D., A. Mas, and J. Rothstein (2008). Tipping and the dynamics of segregation. *The Quarterly Journal of Economics 123*(1), 177–218.

Coleman, J. S. (1958). Relational analysis: the study of social organizations with survey methods. *Human organization 17*(4), 28.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American journal of sociology 94*, S95–S120.

Currarini, S., E. Fumagalli, and F. Panebianco (2017). Peer effects and local congestion in networks. *Games and Economic Behavior 105*, 40–58.

Currarini, S., M. O. Jackson, and P. Pin (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica 77*(4), 1003–1045.

De Martí, J. and Y. Zenou (2017). Segregation in friendship networks. *The Scandinavian Journal of Economics 119*(3), 656–708.

Dev, P. (2014). Identity and fragmentation in networks. *Mathematical Social Sciences 71*, 86–100.

Fagiolo, G., M. Valente, and N. J. Vriend (2007). Segregation in networks. *Journal of Economic Behavior & Organization 64*(3-4), 316–336.

Fagiolo, G., M. Valente, and N. J. Vriend (2009). A dynamic model of segregation in small-world networks. In *Networks, Topology and Dynamics*, pp. 111–126. Springer.

Galeotti, A., S. Goyal, and J. Kamphorst (2006). Network formation with heterogeneous players. *Games and Economic Behavior 54*(2), 353–372.

Goeree, J. K., A. Riedl, and A. Ule (2009). In search of stars: Network formation among heterogeneous agents. *Games and Economic Behavior 67*(2), 445–466.

Golub, B. and Y. Livne (2010). Strategic random networks. *Available at SSRN 1694310*.

Goyal, S. and F. Vega-Redondo (2007). Structural holes in social networks. *Journal of Economic Theory 137*(1), 460–492.

Granovetter, M. (1983). The strength of weak ties: A network theory revisited.

Grauwin, S., F. Goffette-Nagot, and P. Jensen (2012). Dynamic models of residential segregation: An analytical solution. *Journal of Public Economics 96*(1-2), 124–141.

Hostetler, J. A. (1993). *Amish society.* JHU Press.

Iijima, R. and Y. Kamada (2017). Social distance and network structures. *Theoretical Economics 12*(2), 655–689.

Ingram, P. and M. W. Morris (2007). Do people mix at mixers? structure, homophily, and the "life of the party". *Administrative Science Quarterly 52*(4), 558–585.

Jackson, M. O. and B. W. Rogers (2005). The economics of small worlds. *Journal of the European Economic Association 3*(2-3), 617–627.

Jackson, M. O. and A. Wolinsky (1996). A strategic model of social and economic networks. *Journal of economic theory 71*(1), 44–74.

Johnson, C. and R. P. Gilles (2003). Spatial social networks. In *Networks and Groups*, pp. 51–77. Springer.

Kromhout, M. and P. Vedder (1996). Cultural inversion in afro-caribbean children in the netherlands. *Anthropology & education quarterly 27*(4), 568–586.

McPherson, J. M. and L. Smith-Lovin (1986). Sex segregation in voluntary associations. *American Sociological Review*, 61–79.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual review of sociology 27*, 415–444.

Neto, F., J. Barros, and P. G. Schmitz (2005). Acculturation attitudes and adaptation among portuguese immigrants in germany: Integration or separation. *Psychology and Developing Societies 17*(1), 19–32.

Pancs, R. and N. J. Vriend (2007). Schelling's spatial proximity model of segregation revisited. *Journal of Public Economics 91*(1-2), 1–24.

Phinney, J. S., G. Horenczyk, K. Liebkind, and P. Vedder (2001). Ethnic identity, immigration, and well-being: An interactional perspective. *Journal of social issues 57*(3), 493–510.

Ruef, M., H. E. Aldrich, and N. M. Carter (2003). The structure of founding teams: Homophily, strong ties, and isolation among u.s. entrepreneurs. *American sociological review 68*, 195–222.

Safi, S. (2018). *Essays on networks, social ties, and labor markets.*

Schelling, T. C. (1969). Models of segregation. *The American Economic Review 59*(2), 488–493.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of mathematical sociology 1*(2), 143–186.

Shrum, W., N. H. Cheek Jr., and S. MacD (1988). Friendship in school: Gender and racial homophily. *Sociology of Education 61*, 227–239.

Twist, K. A. (2019). *Partnering with Extremists: Coalitions Between Mainstream and Far-right Parties in Western Europe*. University of Michigan Press.

Zhang, J. (2011). Tipping and residential segregation: a unified schelling model. *Journal of Regional Science 51*(1), 167–193.