2017 IEEE 56th Annual Conference on Decision and Control (CDC)
December 12-15, 2017, Melbourne, Australia

Dynamics of Group Cohesion in Homophilic Networks

Katerine Guerrero and Jorge Finke

Abstract— **Understanding cohesion and homophily in empirical networks allows us build better personalization and recommendation systems. This paper proposes a network model that explains the emergence of cohesion and homophily as an aggregate outcome at the group- and network-level. We introduce two simple mechanisms that capture the underlying tendencies of nodes to connect with similar and different others. Our main theoretical result presents conditions on the network under which it reaches high degrees of cohesion and homophily.**

I. INTRODUCTION

Networks are generally composed of different types of nodes. A node may represent a social actor or a piece of information, and each type indicates a distinctive trait, interest, or function. The concept of homophily captures the tendency of nodes to connect with other nodes of the same type (also known in epidemiology as assortative mixing) [1], [2]. A number of studies identify homophily as one of the main determinants of the structure of empirical networks [3].

Homophily is a global property of the network, measured by how nodes of the same type connect to each other in comparison to a random mechanism for establishing links [4]. If the number of links across groups is less than the expected number of links due to the random mechanism (by orders of magnitude), the network is said to exhibit homophily.

Unlike homophily, cohesion is an aggregate measure of a group of nodes of the same type. It represents a degree of membership to a group [5] and different groups hold different values. A group in which members have a high fraction of links to other group members is considered a cohesive group [6].

It is an open challenge to characterize the evolution of group cohesion and homophily. The most common approach is to model local stochastic mechanisms that describe how nodes establish and remove links. A number of models focus on scenarios where the mechanisms depend on the type associated to a node. Mechanism-based models provide a framework to understand cause-effect relationships underlying group cohesion and homophily and to characterize the extent to which different types of nodes contribute to the evolution of particular network structures.

This paper introduces a model that captures the dynamics of the connections between nodes in a network composed by two groups. Decision-making is driven by the assumption that nodes of a particular type obey, with a given probability, two simple mechanisms. These mechanisms encourage or discourage establishing links to similar others. We analyze the dynamics of this model as a stochastic Markov process, and find that transition probabilities of the state of cohesion of a group are a function of the relative size of the group.

The rest of this paper is organized as follows. Section II introduces the proposal model of interaction between nodes. Section III presents the analytical results, which show how the mechanisms impact the cohesion of a group. Finally, Section IV presents some simulation results and identifies conditions under which the network exhibits homophily. The proofs of the lemmas and the theorem are available at http://jfinke.org/research/publications.

K. Guerrero is with the School of Systems and Computer Engineering, Universidad del Valle, Cali, Colombia.

J. Finke is with the Department of Electrical Engineering and Computer Science, Pontificia Universidad Javeriana, Cali, Colombia.

II. THE NETWORK MODEL

Let $N = \{1, \ldots, n\}$, $n \in \mathbb{N}$, represent a set of nodes, which are divided into two groups based on common traits or interest. Nodes within the same group are said to be of the same type. Let $p_i : N \mapsto \{0, 1\}$ denote the type associated to node i. The set $N_\ell = \{i \in N : p_i = \ell\}$, $\ell \in \{0, 1\}$, groups all nodes of type ℓ and $n_\ell = |N_\ell|$ denotes the group size. Let $n_0 \leq n_1$. We refer to N_0 and N_1 as the minority and majority group, respectively.

A directed graph $G = (N, M)$ captures the connections between nodes, where $M = \{m_{ij}\}$, $m_{ij} \in \{0, 1\}$. In particular, $m_{ij} = 1$ if there exists a link from node i to node j. Links may be reciprocal, but there are no self-loops, that is, $m_{ii} = 0$ for all $i \in N$. The set of neighbors of node i, denoted by $Q_i = \{j \in N : m_{ij} = 1\}$, refers to the nodes to which node i establishes outgoing links. The set $Q'_i = \{j \in Q_i : p_i = p_j\}$ refers to the neighbors of the same type. Assume that each node has the same number of neighbors $q = |Q_i|$, but not necessarily the same number of neighbors of the same type, that is, $|Q'_i| \neq |Q'_j|$ for some $i, j \in N$. Moreover, let $2 < q < n_0$. Since nodes of both types have the same number of neighbors, the maximum number of neighbors is bounded by the size of the minority group. Note that establishing more than $n_0 - 1$ links would force the minority nodes to connect to majority nodes (i.e., the minority group could not be totally cohesive).

The state of the network at time t is defined as $x(t) = [x_1(t), \ldots, x_n(t)]^\top$, where each element

$$x_i = \frac{|Q'_i|}{q} \tag{1}$$

represents the proportion of same-type neighbors of node i. The decision by node $i \in N$ to remove and establish links follows two mechanisms:

M1. Node i disconnects from neighboring node $j \in Q_i$ with $p_j \neq p_i$ and connects to node $k \notin Q_i$.

M2. Node i disconnects from neighboring node $j \in Q_i$ and connects to node $k \notin Q_i$ with $p_k \neq p_i$.

Mechanism M1 captures the tendency to remove links to nodes of different type and to replace them with random links to nodes of either type. Mechanism M2, in contrast, captures the tendency to replace links to randomly selected neighbors with links to nodes of a different type. Each mechanism guarantees that the number of (outgoing) neighbors remains constant. Together, M1 and M2 lead to the formation of cohesive relationships between nodes of the same type, without any form of centralized coordination. To characterize the coupling between nodes consider the following definitions.

Definition 1: Node i is said to be type-neutral if $x_i = 1/2$; non-cohesive if $x_i \leq 1/2$.

Note that a type-neutral node establishes half of its links to nodes of the same type (i.e., $|Q'_i| = q/2$). This requires that the total number of neighbors q is even. Note also that the number of links to nodes of the same type of a non-cohesive node satisfies $|Q'_i| \leq q/2$.

Definition 2: The cohesion index of the group of type ℓ is defined as

$$h_\ell = \frac{1}{n_\ell} \sum_{i \in N_\ell} x_i \tag{2}$$

Equation (2) represents the average fraction of connections of nodes of type ℓ with other nodes of the same type[1]. Note that the value of h_ℓ depends only on the neighbors of the nodes of type ℓ. For the set N_ℓ to form a cohesive group, we expect the average number of links to neighbors of the same type to be significantly larger than the average number of links to neighbors of different type.

According to Definitions 1 and 2, if every node $i \in N_\ell$ is type-neutral, then the group N_ℓ must have a cohesion index $h_\ell = 1/2$. However, note that the group N_ℓ may satisfy $h_\ell = 1/2$ and not all nodes of type ℓ are type-neutral.

Definition 3: The group N_ℓ shows total cohesion if $h_\ell = 1$. If $h_\ell = 0$, then there is no group cohesion.

Total cohesion of a group means that all outgoing links of its nodes are directed towards nodes within the same group. If the two groups show total cohesion then the network G is said to be segregated, meaning that each node connects only to nodes of the same type (i.e., $x_i = 1$ for all $i \in N$).

At time t, let $e_i(t)$ denote an event that node i disconnects from a node and connects to another. An event e_i is of type 1 if the event is triggered by mechanism M1. The set $E_1(t)$ denotes all possible events of type 1. These events only occur for nodes with a state strictly less than 1 (if $x_i = 1$, then node i does not have a different-type neighbor from which to disconnect). That is, for every $e_i \in E_1$, the state of node i is $0 \leq x_i \leq (q-1)/q$. Note that events of type 1 tend to disconnect nodes from neighboring nodes of different type and may increase (but not decrease) the cohesion index h_ℓ.

Similarly, the set $E_2(t)$ groups all possible events triggered by mechanism M2. These events occur for a node with any state ($0 \leq x_i \leq 1$). Events of type 2 tend to connect nodes to other nodes of different type and may decrease (but not increase) the cohesion index h_ℓ.

[1]The definition of the index is equivalent to the one introduced by Currarini et al [7].

Let $g_e(x)$ be a function that enables an event at time t. If $e_i \in g_e(x(t))$, the next state of the network is defined by $x(t+1) = f_e(x(t))$, where the operator f_e specifies the state transitions. In particular, if $e_i \in g_e(x(t))$ such that node i disconnects from node j and connects to node k, then

$$x_i(t+1) = \begin{cases} x_i(t) + \dfrac{1}{q}, & \text{if } p_i \neq p_j \text{ and } p_i = p_k; \\ x_i(t) - \dfrac{1}{q}, & \text{if } p_i = p_j \text{ and } p_i \neq p_k; \\ x_i(t), & \text{if } p_i \neq p_j \text{ and } p_i \neq p_k. \end{cases} \tag{3}$$

According to M1 and M2, it is not possible for nodes i, j, and k to be of the same type. The enable function $g_e(x)$ together with the state transition operator $f_e(x)$ define the evolution of the network.

Consider the following assumption.

Assumption 1: If $e_i \in g_e(x)$, then $e_i \in E_1$ with probability ε. Furthermore, node i of type $p_i = \ell$ satisfies:

(a) The cohesion index $h_\ell \in (0, 1)$.

(b) The selection of nodes k and j follows independent uniform distributions.

Assumption 1 considers that events of type 1 occur with probability ε and events of type 2 with probability $1 - \varepsilon$. Condition (a) requires that nodes belonging to a group with either total or no cohesion do not rearrange their links. If at time t' the network satisfies $h_0(t'), h_1(t') \in \{0, 1\}$, then there are no enabled events of either type. We model the deadlock of the network dynamics by defining e_0 such that $\forall t \geq t'$, $e_0 \in g_e(x(t))$, $x_i(t+1) = x_i(t')$ (i.e., the state does not change over time). Condition (b) requires that mechanisms M1 and M2 select which node to connect to and disconnect from based random uniform distributions. Note that establishing new links (or removing existing ones) does not depend on the degree of the nodes or any other measure, except its type. Assumption 1 allows us to study how the set of all trajectories starting from a network in which every node is type-neutral results in strong group cohesion for both groups.

III. Analytical Results

We first characterize the probabilities that node i removes links and replaces them according to mechanisms M1 and M2. Second, we define the transition probabilities of the cohesion index of group N_ℓ and determine the probability that group N_ℓ shows no cohesion before total cohesion.

A. Probabilities of Node Decision-Making

Based to mechanisms M1 and M2, consider the probabilities that node i selects nodes j and k of a particular type. Let $P[p_i \neq p_j, p_i = p_k]$ denote the probability that node j is not but node k is of type ℓ. Moreover, let $P[p_i = p_j, p_i \neq p_k]$ denote the probability that node j is of type ℓ but node k is not. And finally, let

$P[p_i \neq p_j, p_i \neq p_k]$ denote the probability that both nodes j and k are not of type ℓ.

Lemma 1: The probabilities that node i, with $p_i = \ell$, establishes links to other nodes (of the same or different type), given event $e_i \in g_e(x)$, are

$$P[p_i \neq p_j, p_i = p_k] = \begin{cases} \varepsilon \, \dfrac{n_\ell - qx_i - 1}{n - q - 1}, & \text{if } x_i < 1; \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

$$P[p_i = p_j, p_i \neq p_k] = (1 - \varepsilon)\, x_i \tag{5}$$

$$P[p_i \neq p_j, p_i \neq p_k] = \begin{cases} 1 - \varepsilon \, \dfrac{n_\ell - qx_i - 1}{n - q - 1} - (1 - \varepsilon)\, x_i, & \\ \qquad\qquad \text{if } x_i < 1; & \\ 0, \qquad\qquad \text{otherwise.} & \end{cases} \tag{6}$$

Equations (4)-(6) characterize the probability with which the transition operator f_e satisfies each conditional (see (3)). Using (4), if $\varepsilon = 0$ (i.e., if events of type 1 are not enabled), then $P[p_i \neq p_j, p_i = p_k] = 0$. And according to (5), if $\varepsilon = 1$ (i.e., if events of type 2 are not enabled), then $P[p_i = p_j, p_i \neq p_k] = 0$.

Fig. 1 depicts (4) and (5) as a function of x_i. The black and grey lines intersect at $\left(x^\ell, (1-\varepsilon)x^\ell\right)$. Consider $0 < x^\ell < 1$. Note that if node i satisfies $x_i < x^\ell$, then $P[p_i = p_j, p_i \neq p_k] < P[p_i \neq p_j, p_i = p_k]$. That is, the probability that node i disconnects from a node of different type and connects to a node of the same type is higher than the probability that node i disconnects from a node of the same type and connects to a node of different type. Consequently, the state of node i is more likely to increase than to decrease. Note also that if $x_i > x^\ell$, then x_i is more likely to decrease than to increase.

B. Transition Probabilities between Group Cohesion Indices

We apply Lemma 1 to characterize the transition probabilities between cohesion indices for a group. Based on Definition 2 and (3), if $e_i \in g_e(x)$ and $p_i = \ell$, the cohesion of group N_ℓ at time $t+1$ is given by

$$h_\ell(t+1) = \begin{cases} h_\ell(t) + \dfrac{1}{qn_\ell}, & \text{if } p_i \neq p_j \text{ and } p_i = p_k; \\ h_\ell(t) - \dfrac{1}{qn_\ell}, & \text{if } p_i = p_j \text{ and } p_i \neq p_k; \\ h_\ell(t), & \text{if } p_i \neq p_j \text{ and } p_i \neq p_k. \end{cases} \tag{7}$$

Note that updates in the value of group cohesion depend on the size of the group. Now, let $c_\ell = qn_\ell + 1$ and

$$H_\ell = \left[w_1, w_2, \ldots, w_{c_\ell - 1}, w_{c_\ell}\right]^\top \tag{8}$$

be a vector containing all possible values of group cohesion, represented by $w_1, \ldots, w_{c_\ell}$. Let u be the position of $w_u = (u-1)/(qn_\ell)$ in H_ℓ. The *transition probability* between cohesion indices w_u and w_v is defined as $\pi_{uv} = P[h_\ell(t+1) = w_v | h_\ell(t) = w_u]$. Based on (7), we know that $\pi_{uv} = 0$ for $v = u \pm k$, $k > 1$. Next, we specify the transition probabilities for each group such that $u \geq 1$ and $v \in \{u-1,\ u,\ u+1\}$. Note that $\sum_{v \in \{u-1,\ u,\ u+1\}} \pi_{uv} = 1$. In particular, according to Assumption 1(a) if $u \in \{1, c_\ell\}$, then $w_u \in \{0, 1\}$ and no events of type 1 or 2 can occur.

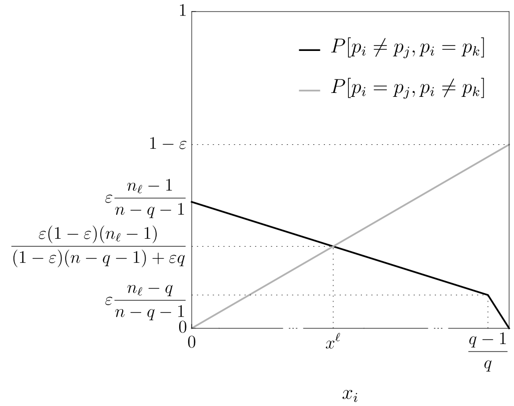

Fig. 1. Probabilities that a node of type ℓ (i.e., node i with $p_i = \ell$) disconnects from and connects to different types of nodes.

Thus, $\pi_{uv} = 1$ for $v = u$ and $\pi_{uv} = 0$ for $v \neq u$. In other words, $\{w_1\}$ and $\{w_{c_\ell}\}$ are closed subsets with no transition to other cohesion indices $\{w_2, \ldots, w_{c_\ell - 1}\}$, meaning that w_1 and w_{c_ℓ} are *absorbing* cohesion indices.

Next, let $T_\ell = \{w_2, \ldots, w_{c_\ell - 1}\}$ denote the set of non-absorbing cohesion indices. Moreover, let the transitions probabilities from the cohesion index $w_u \in T_\ell$ to adjacent cohesion indices (i.e., from w_u to w_v, $v = u \pm 1$) be denoted by $\bar{\pi}_u = \pi_{u(u-1)}$ and $\hat{\pi}_u = \pi_{u(u+1)}$.

Lemma 2: The cohesion index $w_u \in T_\ell$ decreases with probability

$$\bar{\pi}_u = (1-\varepsilon)\, w_u \tag{9}$$

and the probability that the cohesion index w_u increases is bounded by

$$\hat{\pi}_u \;\geq\; \min\{\hat{\pi}_u\} \;=\; \begin{cases} \varepsilon \, \dfrac{n_\ell - qw_u - 1}{n-q-1}, & \\ \qquad \text{if } u \in \{2, \ldots, c_\ell - n_\ell - 1\}; & \\ \varepsilon \, \dfrac{n_\ell - q}{n-q-1}, & \\ \qquad \text{if } u \in \{c_\ell - n_\ell, \ldots, c_\ell - 1\}. & \end{cases} \tag{10}$$

$$\hat{\pi}_u \;\leq\; \max\{\hat{\pi}_u\} \;=\; \varepsilon \, \frac{n_\ell - 1}{n-q-1} \tag{11}$$

According to Lemma 2, the transition probability $\bar{\pi}_u$ depends on both the current cohesion index w_u and the probability ε, while the transition probability $\hat{\pi}_u$ depends on the current index w_u, the probability ε, and the group size n_ℓ. Equations (10) and (11) define bounds on the probability $\hat{\pi}_u$.

Note that the transition probabilities do not depend on time, which allows us to characterize the evolution of the structure of the network as a *homogeneous discrete-time Markov chain* [8]. The transition diagram for the cohesion indices is shown in Fig. 2. If $u \in \{2, \ldots, c_\ell - 1\}$, then the probability $\pi_{uu} = 1 - \hat{\pi}_u - \bar{\pi}_u$. Otherwise, if $u = 1$, then $\hat{\pi}_1 = 0$, and if $u = c_\ell$, then $\bar{\pi}_{c_\ell} = 0$.

Remark 1: Using (10) and (11), because $q > 2$, we know that for any index $w_u \in T_\ell$, $\min\{\hat{\pi}_u\} < \max\{\hat{\pi}_u\}$ holds with strict inequality.

Next, we show how the transition probabilities defined in Lemma 2 evolve with respect to the current cohesion index w_u. First, consider the probability $\bar{\pi}_u$ (defined by (9)). It can be shown that if $\varepsilon < 1$, then $0 < \bar{\pi}_2 < \cdots < \bar{\pi}_{c_\ell - 1} < 1 - \varepsilon$. That is, $\bar{\pi}_u$ is strictly increasing over T_ℓ (see the grey curve in Fig. 3). Second, consider the lower bound on the probability $\hat{\pi}_u$ (defined by (10)). It can be shown that if $\varepsilon > 0$, then $\min\{\hat{\pi}_2\} > \cdots > \min\{\hat{\pi}_{c_\ell - n_\ell}\} = \cdots = \min\{\hat{\pi}_{c_\ell - 1}\}$. That is, $\min\{\hat{\pi}_u\}$ is decreasing over T_ℓ (see the black solid curve in Fig. 3). Third, consider the upper bound on the probability $\hat{\pi}_u$ (defined by (11)). Note that $\max\{\hat{\pi}_u\}$ does not depend on the current cohesion index. Therefore, we know that $\max\{\hat{\pi}_u\}$ is constant (see the black dashed line in Fig. 3).

C. Markov Chain Analysis

We use the transition probabilities of Lemma 2 to obtain the probability that the Markov chain visits one of the closed set before the other (based on a similar argument as in [9, p. 386]). The probability that a chain starting at $w_u \in T_\ell$ reaches the cohesion index w_1 before w_{c_ℓ} is denoted by $\alpha_u = P[t_{u1} < t_{uc_\ell}]$, where $t_{u1} = \min\{t > 0 : h_\ell(0) = w_u, h_\ell(t) = w_1\}$ and $t_{uc_\ell} = \min\{t > 0 : h_\ell(0) = w_u, h_\ell(t) = w_{c_\ell}\}$. Note that $\alpha_1 = 1$ and $\alpha_{c_\ell} = 0$.

Theorem 1: The probability that group N_ℓ, starting at $w_u \in T_\ell$, shows no cohesion before total cohesion is

$$\alpha_u = \prod_{k=2}^{u} \left(1 - \frac{1}{1 + \sum_{i=k}^{c_\ell - 1} \prod_{j=k}^{i} \frac{\bar{\pi}_j}{\hat{\pi}_j}} \right) \tag{12}$$

Theorem 1 characterizes the probability that the first time a chain enters w_1 is less than the first time it enters w_{c_ℓ}, starting from $w_u \in T_\ell$. Due to the bounds on the probability that the cohesion index $w_u \in T_\ell$ increases ((10) and (11)), (12) is also bounded. For the group N_ℓ and $u \in \{2, \ldots, c_\ell - 1\}$ we have that

$$\alpha_u \geq \min\{\alpha_u\} = \prod_{k=2}^{u} \left(1 - \frac{1}{1 + \sum_{i=k}^{c_\ell - 1} \prod_{j=k}^{i} \frac{\bar{\pi}_u}{\max\{\hat{\pi}_u\}}} \right)$$

$$\alpha_u \leq \max\{\alpha_u\} = \prod_{k=2}^{u} \left(1 - \frac{1}{1 + \sum_{i=k}^{c_\ell - 1} \prod_{j=k}^{i} \frac{\bar{\pi}_u}{\min\{\hat{\pi}_u\}}} \right)$$

Based on Remark 1, we know that for any index $w_u \in T_\ell$, $\min\{\alpha_u\} < \max\{\alpha_u\}$ holds with strict inequality. Moreover, it can be shown that if $0 < \varepsilon < 1$, then $1 > \min\{\alpha_2\} > \cdots > \min\{\alpha_{c_\ell - 1}\} > 0$ and $1 > \max\{\alpha_2\} > \cdots > \max\{\alpha_{c_\ell - 1}\} > 0$. That is, the bounds on α_u are strictly decreasing. Fig. 4 evaluates the upper and lower bounds of α_u as a function of w_u. The shaded region represents the difference between the two bounds.

IV. Simulation Results

Next, we consider different scenarios for the evolution of the network. In the first scenario, suppose that Assumption 1 holds. We compare theoretical and simulation results of the

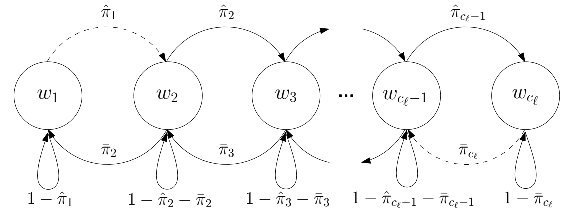

Fig. 2. Transition diagram for the cohesion indices for group N_ℓ (under Assumption 1, $\hat{\pi}_1 = \bar{\pi}_{c_\ell} = 0$).

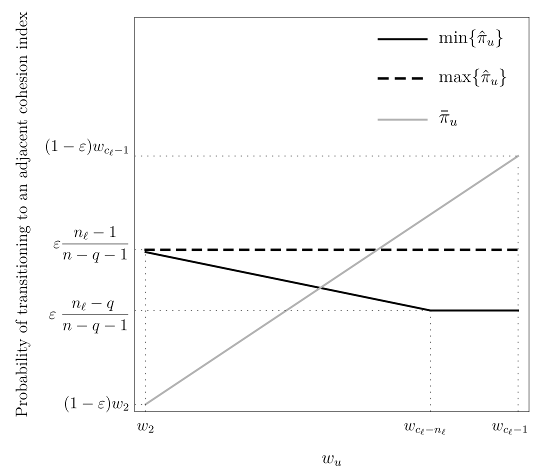

Fig. 3. Probabilities that the cohesion index $w_u \in T_\ell$ for group N_ℓ increases or decreases.

transition probabilities of the cohesion index of each group (defined by (9)-(11)). Let $q = 4$, $n_0 = 10$, and $n_1 = 50$. Figs. 5-8 show the estimated values and the theoretical predictions for $\varepsilon = 1/2$. In particular, Figs. 5 and 6 give insight into the majority group. The box plots represent the estimates of the probabilities $\bar{\pi}_u$ and $\hat{\pi}_u$ as a function of of w_u, for $u \in \{(c_1 + 1)/2, \ldots, c_1 - 1\}$. Figs. 7 and 8 illustrate the estimates of the probabilities $\bar{\pi}_u$ and $\hat{\pi}_u$ as a function of w_u, for $u \in \{2, \ldots, (c_0 + 1)/2\}$, for the minority group. For both groups, the theoretical value coincides with the average estimate of the probability $\bar{\pi}_u$ (based on 50 simulation runs). Moreover, all estimates of $\hat{\pi}_u$ are within the bounds established by (10) and (11).

In the second scenario, suppose that nodes that belong to a group showing either total or no cohesion may rearrange their links.

Assumption 2: If $e_i \in g_e(x)$, then $e_i \in E_1$ with probability ε. Furthermore, node i of type $p_i = \ell$ satisfies:

a. *The cohesion of the group of node i lies in the range $0 \leq h_\ell \leq 1$.*
b. *Nodes k and j are selected based on independent random uniform distributions.*

Note that Assumption 2 relaxes Assumption 1. Under Assumption 2, the cohesion indices representing total and no cohesion are no longer absorbing indices. Let $G(0)$ be a network in which every node is type-neutral and suppose that Assumption 2 holds. First, consider a network with $q = 4$, $n_1 = 50$ and $n_0 \in \{5, 10, \ldots, 50\}$. Let $\varepsilon = 1/2$. Fig. 9 shows the average cohesion indices for the

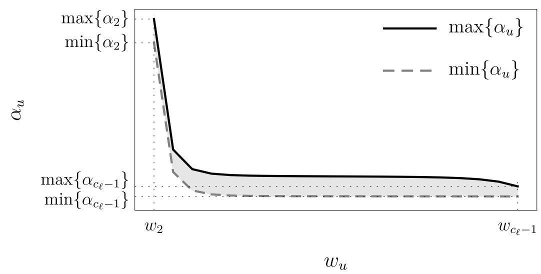

Fig. 4. Bounds on α_u as a function of w_u.

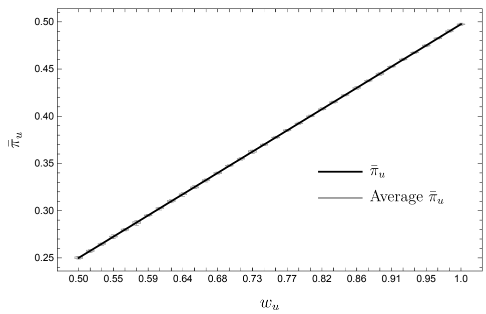

Fig. 5. Probability that cohesion index for the majority group decreases as a function of the current cohesion index $w_u \in T_1$.

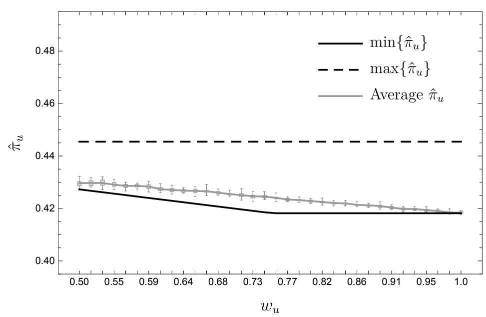

Fig. 6. Probability that cohesion index for the majority group increases as a function of the current cohesion index $w_u \in T_1$.

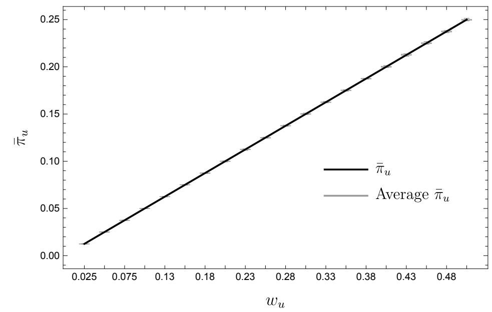

Fig. 7. Probability that cohesion index for the minority group decreases as a function of the current cohesion index $w_u \in T_0$.

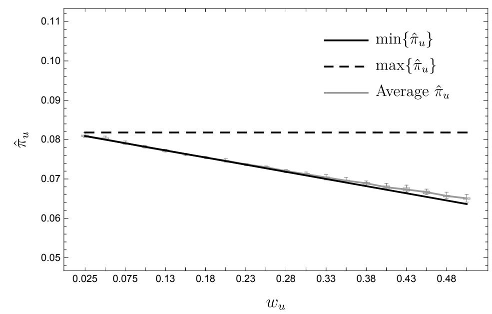

Fig. 8. Probability that cohesion index for the minority group increases as a function of the current cohesion index $w_u \in T_0$.

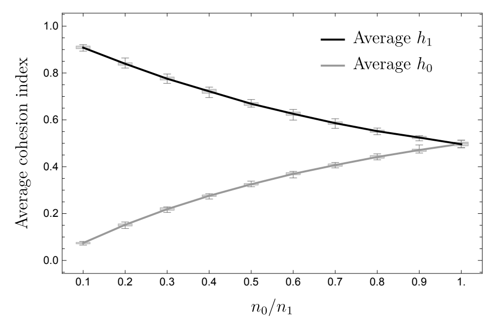

Fig. 9. Average cohesion index for group N_ℓ.

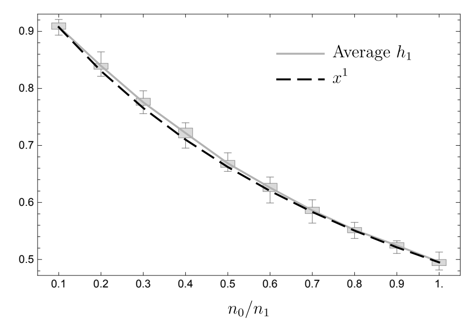

Fig. 10. Average cohesion index for the majority group and x^1.

majority and minority nodes as a function of group size. For any value $n_0/n_1 < 1$ the majority group reaches an average cohesion index above $1/2$. Moreover, the greater the ratio n_0/n_1, the smaller the average cohesion index of the majority group. The opposite is true for the minority group. Fig. 10 (Fig. 11) shows the average cohesion index and the value of x^ℓ for the majority group (for the minority group, respectively). Remember that x^ℓ refers to the intersection between $P[p_i \neq p_j, p_i = p_k]$ (defined by (4)), and $P[p_i = p_j, p_i \neq p_k]$ (defined by (5)). The small difference between the average cohesion index and x^ℓ suggests a direct relationship between the probability of connecting to a node of a particular type and the resulting cohesion of the group.

Second, consider $q = 4$, $n_1 = 50$ and $n_0 = 10$. Fig. 12 shows the average cohesion indices for both groups as a function of ε. The higher the value of ε, the stronger the average group cohesion. Note that if $0 < \varepsilon < 1$, then the average cohesion index for the minority group is always less than for the majority group. Note also that if $\varepsilon = 0$, then event $e_i \in g_e(x)$ is always of type 2. That is, the state of node i may decrease, but not increase. Moreover, if node i reaches the state $x_i = 0$, then the set the neighbors of node i remains fixed (because the probability that node i redirects a link to other node of its group towards a node of the other group is $P[p_j = p_i, p_k \neq p_i] = 0$). Therefore, according to (2), if $\varepsilon = 0$ then the cohesion index h_ℓ may decrease, but not increase. Moreover, if $x_i = 0$ for all $i \in N_\ell$, then group N_ℓ shows no cohesion (i.e., $h_\ell = w_1 = 0$) and the

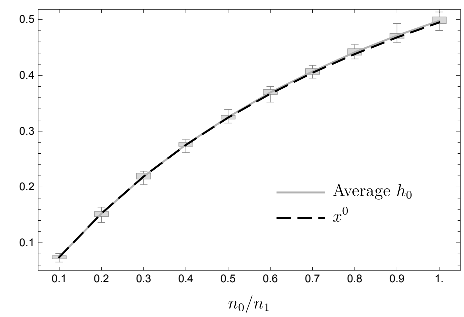

Fig. 11. Average cohesion index for the minority group and x^0.

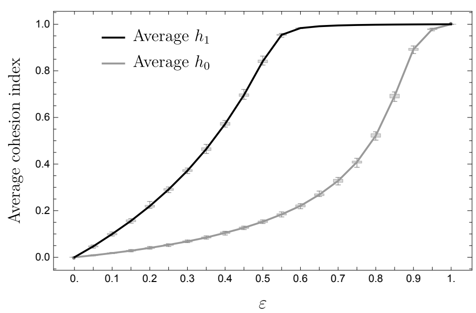

Fig. 12. Average cohesion index of the type ℓ as function of ε.

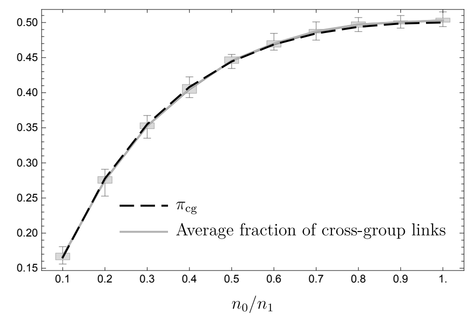

Fig. 13. Average fraction of cross-group links when $\varepsilon = 1/2$.

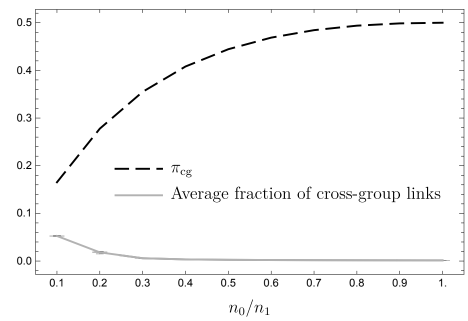

Fig. 14. Average fraction of the cross-group links when $\varepsilon = 9/10$.

cohesion index w_1 is an absorbing index. Similarly, if $\varepsilon = 1$, then event $e_i \in g_e(x)$ is always of type 1. That is, the state of node i may increase, but not decrease. Moreover, if node i reaches the state $x_i = 1$, then the set of neighbors of node i remains fixed (because the probability that a node i redirects a link to other node of other group towards a node of its group is $P[p_j \neq p_i, p_k = p_i] = 0$). Therefore, according to (2), if $\varepsilon = 1$ then the cohesion index h_ℓ may increase, but not decrease. Moreover, if $x_i = 1$ for all $i \in N_\ell$, then the group N_ℓ shows total cohesion (i.e., $h_\ell = w_{c_\ell} = 1$) and the cohesion index w_{c_ℓ} is an absorbing index.

Third, we follow a similar argument as in [4, p. 88] to evaluate whether the network exhibits homophily based on the two types of nodes. A network shows no homophily when nodes of different types establish connections regardless of their type. Therefore, for a network with no homophily the probability that the first end of a given link is to a node of type p_i and the second end to a node of type $p_j \neq p_i$ (or viceversa) is

$$\pi_{\text{cg}} = \frac{2n_0 n_1}{n^2}$$

where n_ℓ/n is the probability that node i belongs to the group N_ℓ. If the fraction of cross-group links is significantly less than π_{cg}, then the network exhibits homophily. Consider a network with $q = 4$, $n_1 = 50$ and $n_0 \in \{5, 10, \ldots, 50\}$. Fig. 13 shows the values of π_{cg} and the average fraction of cross-group links of the simulated network when $\varepsilon = 1/2$. Note that the difference between the two curves is insignificant, which suggests that if $\varepsilon = 1/2$, then the network shows no homophily regardless of the relative size difference between groups. Fig. 14 illustrates the resulting homophily when $\varepsilon = 9/10$. Note that the average fraction of cross-group links is significantly less than π_{cg}.

ACKNOWLEDGMENTS

This research was supported in part by the Center of Excellence and Appropriation in Big Data and Data Analytics (CAOBA) at the Pontificia Universidad Javeriana, the Ministry of Information Technologies and Telecommunications of Colombia (MinTIC), and the Colombian Administrative Department of Science, Technology and Innovation (COLCIENCIAS), under grant no. FP44842-546-2015.

REFERENCES

[1] M. McPherson, L. Smith-Lovin, and J. Cook, "Birds of a feather: Homophily in social networks," *Annual Review of Sociology*, vol. 27, no. 1, pp. 415–444, 2001.

[2] M. E. J. Newman, "Mixing patterns in networks," *Physical Review E*, vol. 67, no. 2, p. 026126, 2003.

[3] V. Ciotti, M. Bonaventura, V. Nicosia, P. Panzarasa, and V. Latora, "Homophily and missing links in citation networks," *EPJ Data Science*, vol. 5, no. 1, pp. 1–14, 2016.

[4] D. Easley and J. Kleinberg, *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.

[5] N. E. Friedkin, "Social cohesion," *Annual review of sociology*, vol. 30, pp. 409–425, 2004.

[6] R. D. Alba, "A graph-theoretic definition of a sociometric clique," *Journal of Mathematical Sociology*, vol. 3, no. 1, pp. 113–126, 1973.

[7] S. Currarini, M. O. Jackson, and P. Pin, "An economic model of friendship: Homophily, minorities, and segregation," *Econometrica*, vol. 77, no. 4, pp. 1003–1045, 2009.

[8] A. Tornambè, *Discrete-event system theory: an introduction*. World Scientific Publishing, 1995.

[9] C. Cassandras and S. Lafortune, *Introduction to discrete event systems*, 2nd ed. Springer, 2008.