Group Formation in Signed Networks

Enrico Maria Fenooalte† and Fanyuan Meng‡

Physics Department, University of Fribourg, Chemin du Musée 3, 1700 Fribourg, Switzerland
(Dated: July 16, 2021)

Why do some social groups or organizations, initially composed of a few people with the same interests, disintegrate when increasing the number of members? To answer this question, we propose a stochastic model where group members and their relationships are described by a growing signed network with positive and negative signs on both nodes and links. We investigate, both analytically and numerically, the behavior of group cohesion, defined as the fraction of nodes with the same sign, when the network grows with noisy information on the link signs. Our findings suggest strategies to prevent or boost group cohesion in different contexts.

Studying the dynamics of group formation is fundamental to understand, for example, how groups of friends, political parties, or criminal organizations evolve. In recent years, the formation of social groups has aroused the interest of not only sociologists but also physicists who study networks [1–7]. Most studies focus more on the connection structure between group members and consider all nodes and links of the same types [8–10].

However, individuals can be different and both positive and negative relationships exist, such as friendships/hostility or trust/distrust, and traditional network structure cannot express these relationships effectively [11, 12]. So, here we shall consider signed networks, that can capture both positive and negative relationships: a positive link between two nodes indicates that there is affinity, while a negative link indicates hostility. Moreover, we consider also positive and negative nodes [13–15]: two nodes of the same sign represent like-minded individuals; otherwise, they are dissimilar.

In addition, in most group formation models, the existing group members have no decisional power over new individuals, i.e. a single node joins the network maximizing its payoff and the other nodes must accept its decision [16–18]. However, many groups grow through a recruitment process where the existing group members, more or less voluntarily, decide who can join and who cannot [19, 20]. Depending on the social context, the recruitment processes can vary: in a group of friends, a new individual can enter even if he/she is accepted by only one member of the existing group; in more formal institutions, such as the European Union, a candidate is accepted only if he/she is welcome by all the other members.

In this letter, we introduce a stochastic model based on signed networks to study these processes. Since in sociology a social group is defined as a group of people that share common interests [21, 22], we shall focus on the group cohesion, that is the fraction of nodes with the same signs.

Unlike other approaches, such as the voter model and its extensions [23–25], we consider a growing network in which the sign of a node is not affected by other nodes but fixed in time. The only factor that can affect the fraction of positive or negative nodes is the recruitment process.

In our model, starting from a few positive nodes, at each time step, a new candidate node tries to join the group and its admission depends on the links between it and the existing nodes. In general, a positive link means that the candidate is welcome in the group, otherwise it is undesirable.

As in the classic theory of structural balance [26–28], the fundamental assumption is that link signs are related to the node signs: two nodes of the same sign are connected by a positive link, otherwise by a negative link. We shall refer to this as the balance rule.

In the real world, however, it is often difficult to establish whether an individual is suitable for a group. For example, there could be partial or imperfect information on the candidate member. Hence, we assume that the balance rule can be violated, i.e. there is some "noise", that represents the probability of having misleading links between nodes. The possibility of violating the balance rule in signed networks was already explored in [29, 30], even though in a different context, i.e. opinion formation dynamics.

Here, we study analytically and numerically the relationship between the amount of noise in the system and the cohesion of the final group for different recruitment processes and network growth mechanisms.

More formally, at time \( t = 0 \), consider a complete network with \( N_F \) positive nodes and positive links among them. We shall refer to these nodes as founder nodes. At each time step \( t \in \{1, 2, \ldots\} \), a candidate node \( c_t \) tries to join the network and connects to \( m \) random existing nodes. The sign of \( c_t \) \((+1 \text{ or } -1)\) is randomly drawn with equal probability. Moreover, given a certain amount of noise \( \eta \in [0, 0.5] \), the sign of each new link follows the balance rule with probability \( 1 - \eta \); if \( c_t \) and its new neighbor have the same sign, then the link between them is positive with probability \( 1 - \eta \) and negative with probability \( \eta \); if they have opposite signs, the link is negative with probability \( 1 - \eta \) and positive with probability \( \eta \).

Hence, when \( \eta = 0 \), the link signs are consistent with the node signs; when \( \eta = 1/2 \), there is no correlation between the link and node signs.
As previously stated, the sign of the link with \( c_t \) represents the willingness to admit (positive link) or reject (negative link) it into the network. Depending on which value of \( m \) we choose, we can distinguish two types of recruitment rules.

(i) \( m = 1 \) for any \( t > 0 \). In this case, \( c_t \) connects with only one random existing node: if their link is positive, the candidate is accepted into the network; otherwise, it is rejected. Since, at each time step, a single node decides whether \( c_t \) can join or not, without considering the opinion of other nodes, we shall refer to this case as the anarchy rule (AR).

(ii) \( m = pN_t \), where \( N_t \) is the number of nodes in the network at time \( t \), and \( p \in (0,1] \) is the connection probability, i.e. we assume that each existing node connects to \( c_t \) with probability \( p \) (if no link is formed, a new candidate node \( c_t \) is drawn). In this case, \( c_t \) connects, on average, to a fraction \( p \) of the existing nodes: if a strict majority of the newly-formed links are positives, the candidate is admitted into the network; otherwise, it is rejected. Since the fate of \( c_t \) is decided democratically by a fraction of the existing nodes, we shall refer to this case as the democracy rule (DR).

Finally, independently of the recruitment rule that we use, the process stops when the network reaches the desired size \( N > N_F \). Note that, in general, \( t \geq N_t - N_F \) and, at the end of the process, the number of accepted candidate nodes is \( N - N_F \).

Here, since we consider positive founder nodes, we define the group cohesion \( C \) as the fraction of positive nodes in the network excluding the founder nodes, i.e.

\[
C = \frac{N^+ - N_F}{N - N_F}, \tag{1}
\]

where \( N^+ \) is the number of positive nodes.

In the following, we will study the behavior of \( C \) as \( \eta \), \( N \), and \( N_F \) vary, both for the AR and the DR.

Let us first consider the AR. To study the behavior of cohesion analytically, we can use a master equation approach and calculate the probability distribution \( P(N^+,N) \) that the number of positive nodes is \( N^+ \) when there are \( N \) nodes in the network (see SI). However, here we only focus on the expected value of the cohesion, \( E_C(N,N_F,\eta) \). Noting that, when the noise is \( \eta \), the transition probability \( P(N^+ \to N^+ + 1|N) \), that a positive candidate node is accepted when there are \( N^+ \) positive nodes out of \( N \), is

\[
P(N^+ \to N^+ + 1|N) = (1 - \eta) \frac{N^+}{N} + \eta \frac{N - N^+}{N}, \tag{2}
\]

then the expected number of positive nodes when the noise is \( \eta \) and there are \( N \) nodes, \( E_{N^+}(N,\eta) \), is given by the following recursive equation:

\[
E_{N^+}(N,\eta) = \frac{N - 2\eta}{N - 1} C(N - 1,\eta) + \eta. \tag{3}
\]

By solving Eq. (3) with the initial condition \( E_{N^+}(N_F,\eta) = N_F \), subtracting the solution by \( N_F \), and dividing it by \( N - N_F \), the expected cohesion is

\[
E_C(N,N_F,\eta) = \frac{N - 2N_F + \frac{\Gamma(N_F + 1)\Gamma(N + 2\eta)}{\Gamma(N + 1)\Gamma(N + 2\eta + 1)(N - N_F)}}{2(N - N_F)}. \tag{4}
\]

Note that, when \( \eta = 1/2 \), we have \( E_C(N,N_F,1/2) = 1/2 \), meaning that, since the node signs are independent of the link signs, the number of positive and negative nodes in the network (excluding the founder nodes) is the same; when \( \eta = 0 \), instead, we have \( E_C(N,N_F,1) = 1 \), meaning that all the nodes are positive since the founder nodes are positive and there are no misleading links.

However, if the noise slightly increases, the expected cohesion drastically decreases (as shown in the top-left panel of Fig. 1). For example, when \( N_F = 1 \) and \( \eta = 0.05 \), the expected cohesion is only about 0.8. This means that even a small amount of noise is enough to damage the group cohesion.

Then, as expected, with the same \( \eta \) and \( N \), the larger \( N_F \) the larger the cohesion (as shown in top-right panel of Fig. 1).

Moreover, when \( N \to \infty \) and \( N_F \) is fixed, we obtain the following relation:

\[
\left[ E_C(N,N_F,\eta) - \frac{1}{2} \right] \propto N^{-2\eta}. \tag{5}
\]

Eq. (5) implies that as the network size increases while keeping \( N_F \) fixed, the expected cohesion decreases to 1/2, independently of the amount of noise. In other words, too large groups cannot be cohesive.

Then, we want to see how the expected cohesion varies when varying the number of founder nodes \( N_F \). If we focus on the ratio \( f_F = N_F/N \), in the large \( N \) limit (hence also large \( N_F \)), we have

\[
\lim_{N \to \infty} E_C(f_F,\eta) = \frac{1 - 2f_F + f_F^{2\eta}}{2(1 - f_F)}. \tag{6}
\]

From Eq. (6) we see that, when the ratio \( f_F \) is fixed, the expected cohesion approaches a constant value, for \( N \to \infty \). This means that an appropriate number of founder nodes can prevent the group from being poorly cohesive due to its large size.

Since, so far, at each time step each existing node has the same probability to connect with \( c_t \), we shall refer to this as the random connection mechanism (RC). Now, since this is a growing network model, it is natural to study the AR variant where the growth process is driven by the preferential attachment mechanism (PA) [31]. Hence, we consider the case where the probability \( P(c_t \to i) \), that an existing node \( i \) connects with \( c_t \), is proportional to the number of its links \( k_i \) [31]:

\[
P(c_t \to i) = \frac{k_i}{\sum_{j=1}^{N} k_j}. \tag{7}
\]
In this case, it is not possible to write a transition probability as in Eq. 2 as now the connection probability depends explicitly on the age of the existing nodes, i.e. on the chronological order that they joined the network.

If we set, for simplicity, \(N_F = 1\), it is possible to show that the probability \(P(\theta_{i+1}^+)\), that the \((i + 1)\)-th node that join the network is positive, is given by the following recursive equation (see SI):

\[
P(\theta_{i+1}^+) = \left(1 - \frac{\eta}{i}\right)P(\theta_i^+) + \frac{\eta}{2i}.
\]

(8)

Since the founder node is positive and, by convention, it is the 0-th node that joined the network, we can solve Eq. 8 with the initial condition \(P(\theta_1^+) = 1\). It is straightforward to generalize this procedure for \(N_F > 1\). Finally, we can write the PA expected cohesion as:

\[
E_C(N, N_F, \eta) = \sum_{i=1}^{N-N_F} P(\theta_i^+) .
\]

(9)

From the explicit solution of Eq. 9 (see SI), we see that also the PA expected cohesion decreases rapidly with noise, and increases with the number of founder nodes (as shown in the top panels of Fig. 1).

Moreover, when \(N \to \infty\) and \(N_F\) is fixed, we have

\[
\left[ E_C(N, N_F, \eta) - \frac{1}{2} \right] \propto N^{-\eta} ,
\]

(10)

implying that, in the large \(N\) limit (with fixed \(N_F\)), also the PA expected cohesion tends to 1/2 independently of the noise.

However, comparing Eq. 10 and Eq. 5, we see that the PA expected cohesion decreases slower with \(\eta\) than the RC one, as shown in the bottom-left panel of Fig. 1. Indeed, the scaling exponent of the PA is two times the RC one. Moreover, the PA expected cohesion is always higher than the RC one for any \(\eta\) and \(N_F\), as shown in the top panel of Fig. 1.

This can be explained intuitively in the following way. Suppose that, when there are \(N\) nodes, the number of positive nodes is \(N^+\). In the RC case, the probability that \(c_i\) connects to a positive node is \(N^+/N\). In the PA case, instead, this probability is higher: indeed, \(c_i\) connects with a higher probability to the nodes that joined earlier the network (including the founder nodes). These, in turn, are more likely to be positive than the nodes that joined later. In other words, it is like the candidate node interacts with a smaller portion of the network, where the cohesion is larger than \(N^+/N\). Hence, if \(\eta \in (0, 1/2)\), the probability that a positive \(c_i\) is accepted into the network is higher in the PA case.

Then, we focus again on the ratio \(f_F\). When keeping \(f_F\) fixed, the PA expected cohesion increases with \(N\) (and hence with \(N_F\)), and it approaches a well-defined limit:

\[
\lim_{N \to \infty} E_C(f_F, \eta) = 1 - \eta.
\]

(11)

Note that, differently from the RC case, Eq. 11 does not depend on \(f_F\), and it is the same expected cohesion that one would obtain in a system where \(c_i\) can only connect with founder nodes. Indeed, the probability that a founder node accepts a positive \(c_i\) is, by definition, exactly \(1 - \eta\). This result was expected: when there is preferential attachment, in the large \(N_F\) limit, the probability of connecting with a node that is not a founder node is negligible.

Moreover, the limit in Eq. 11 is always larger than the one in Eq. 6. To show this, it is enough to note that, when \(f_F \to 1\), Eq. 6 goes to \(1 - \eta\).

Fig. 1 summarizes all these results and shows that our analytical solutions are consistent with numerical simulations.

![Graphs showing mean cohesion](image)

**FIG. 1.** Mean cohesion \(\overline{C}\) for the AR, both in the RC and PA cases. The lines are the analytical solutions in Eq. 6 (blue line) and Eq. 9 (green line). Each point is averaged over 1000 realizations, and the error bars represent twice the standard error of the mean. Top-left: \(\overline{C}\) against \(\eta\), with \(N_F = 1\) and \(N = 100\) fixed. Top-right: \(\overline{C}\) against \(N_F\), with \(N = 100\) and \(\eta = 0.1\) fixed. Bottom-left: \(\overline{C}\) against \(N\), in a log-log scale with \(N_F = 1\) and \(\eta = 0.1\) fixed. Bottom-right: \(\overline{C}\) against \(N_F\), in a log scale on Y axis, with \(f_F = 0.1\) and \(\eta = 0.1\) fixed.

Now, let us consider the DR. We shall show only the numerical results, leaving the analytical solutions as an open problem. Furthermore, we only consider the RC case, as there is no straightforward way to implement the preferential attachment in the DR (see SI).

Recall that \(c_i\) connects to each existing node with probability \(p\). In other words, there are, on average, \(m = pN\) voters that decide democratically whether \(c_i\) can be admitted. If a strict majority of these votes are positives, then \(c_i\) can join the network, otherwise it is rejected.

In Fig. 2, we perform numerical simulations to show how the mean cohesion \(\overline{C}\) (averaged over the realizations) varies with \(\eta\), for different values of \(p\) (for the behavior of \(\overline{C}\) when varying \(N\) and \(N_F\), see SI).

Moreover, we compare these results with the AR in
Eq. 3 and with the case where $c_t$ can only connect with founder nodes. As mentioned before, the expected cohesion of the latter case goes as $1 - \eta$. Since decisions are always made by the same few nodes, we shall refer to this case the *dictatorship rule* (DSR).

From Fig. 2 we note that, for any $p$ and $\eta$, the mean cohesion of the DR is always larger than that of the AR.

Then, the larger the $p$, the larger the cohesion. This means that, as expected, more voters can compensate and prevent individual misjudgments due to the presence of noise. Note that, if $p = 1$, all the existing nodes vote.

However, taking the DSR as a benchmark, we see something interesting. First, there is a critical value $p = p_c$ such that, for $p < p_c$, the cohesion of the DR is smaller than that of the DSR, for any $\eta$ (in the figure, $p_c \approx 0.2$). Second, when $p > p_c$, there is also a critical noise $\eta = \eta_c$: for $\eta < \eta_c$, the cohesion of the DR is larger than that of the DSR; for $\eta > \eta_c$, the DR has a smaller cohesion than the DSR.

Moreover, $\eta_c$ grows with $p$. For example, the figure shows that for $p = 0.6$, we have $0.3 < \eta_c < 0.4$; while for $p = 1$, we have $0.4 < \eta_c < 0.5$.

These results suggest that a democratic group, in which members decide who can join by voting, is more cohesive than a group with a dictatorial system, provided that there are enough voters ($p > p_c$). Nevertheless, when there is too much noise in the system ($\eta > \eta_c$), a dictatorship is preferable to keep the group more cohesive.

To conclude, we introduced a simple model based on signed networks to describe the dynamics of the group cohesion.

Our model assumes that group members can have two states, positive or negative. Of course, the world is more complex, and the reasons that make a group cohesive or disjoint go beyond the above binary assumption. Nevertheless, we capture insightful features. For example, we showed that groups that follow the AR, where each member can decide who can join without consultations with others, are highly sensitive to the noise.

Here we speculate that the AR is more suitable to describe the dynamics of unstructured or informal groups, such as groups of friends. On the other hand, more formal groups, such as political parties or companies, where candidates are judged by a commission composed of several members, are better modeled by the DR. In this case, we showed that the higher the number of members in the commission, the better for building a group of people focused on the same goal.

However, when the information on the candidates is too low (large noise), it is better to give decision-making power to a few members (as in the DSR). We again speculate that this is the reason why, in periods of crisis or chaos (i.e. with a high level of noise), many governments centralize power on some individuals.

We partially solved our model, but an analytical solution for the DR is still missing and will be the goal of future researches. Moreover, our framework opens the possibility to many extensions: for example, one can study other recruitment rules; or one can consider more types of nodes (not only positive or negative).

We believe that research in this direction can help to understand how cohesive groups appear and how they disintegrate, suggesting strategies to policy-makers to create solid groups (for example in government institutions) or avoid their formation (in case of criminal or extremist organizations).

**ACKNOWLEDGMENT**

The authors would like to thank Matus Medo, Berno Buechel, Marco Alexander Schwarz, Manuel Sebastian Mariani and Shizhi Huang who improved this work with relevant discussions. This work was partially supported by the Swiss National Science Foundation (grant no. 200020_182498/1). Fanyuan Meng thanks the China Scholarship Council for financial support.
