MDPI

Article

What is the Entropy of a Social Organization?

Christian Zingg *, Giona Casiraghi *, Giacomo Vaccario * and Frank Schweitzer *

Chair of Systems Design, ETH Zurich, Weinbergstrasse 58, 8092 Zurich, Switzerland
* Correspondence: czingg@ethz.ch (C.Z.); gcasiraghi@ethz.ch (G.C.); gvaccario@ethz.ch (G,V.); fschweitzer@ethz.ch (F.S.)

Version September 10, 2019 submitted to Entropy

Abstract: We quantify a social organization's potentiality, that is its ability to attain different configurations. The organization is represented as a network in which nodes correspond to individuals and (multi-)edges to their multiple interactions. Attainable configurations are treated as realizations from a network ensemble. To have the ability to encode interaction preferences, we choose the generalized hypergeometric ensemble of random graphs, which is described by a closed-form probability distribution. From this distribution we calculate Shannon entropy as a measure of potentiality. This allows us to compare different organizations as well as different stages in the development of a given organization. The feasibility of the approach is demonstrated using data from 3 empirical and 2 synthetic systems.

Keywords: Multi-Edge network; Network Ensemble; Shannon Entropy; Social Organization

1. Introduction

Social organizations are ubiquitous in our everyday life, ranging from project teams, e.g. to produce open source software [1] to special interest groups, such as sports clubs [2] or conference audiences [3] discussed later in this paper. Our experience tells us that social organizations are highly dynamic. Individuals continuously enter and exit, and their interactions change over time. Characteristics like these make social organizations complex and difficult to quantify.

Network science allows to study such complex systems in terms of networks, where nodes represent individuals and edges their interactions [4,5]. Under this assumption, a social organization can be represented by a **network ensemble**. Every network in this ensemble corresponds to one possible configuration of interactions in this organization. Thus, the network that we can reconstruct from observed interactions is only one particular realization from this ensemble. Other configurations can also be realized with a given probability. To cope with this, we need a probability distribution that characterizes the network ensemble and reflects its constraints, such as given numbers of nodes or interactions, or preferences for interactions. The probability space defined by this distribution can then be seen as a way to quantify the *number* and the *diversity* of possible states of a social organization. We argue that such possible states give an indication of the *potentiality* of this organization, i.e., its ability to attain different configurations.

But how can the potentiality be measured? First, we need to decide about a probability distribution suitable for reflecting the interactions and constraints in social organizations. Second, based on this distribution we need to quantify the diversity of the organization. To solve the first task, in this paper we utilize the hypergeometric ensemble, as explained in Section 3.1. To solve the second task, we compute the **information entropy** of this ensemble, as shown in Section 3.2.

Information entropy [6] has recently gained popularity in the study of social organizations. Shannon entropy has, for example, been applied to study communication networks [7], human contact events [8], or anomalies in computer networks on a campus [9]. By generalizing the concept of entropy,

even complexity measures [10] or classifications for complex systems [11] have been suggested. Finally, entropies have also been applied in combination with network ensembles to analyze complexity behind structural constraints [12] or spatial constraints [13] or how restrictive topological constraints are [14].

Recent works that combine network ensembles and entropy analyze the effect of *predefined* constraints. For example, in [14], the author studies how entropy changes when fixing degree sequences or community structures, that are derived from the network topology. By enforcing such topological constraints, the resulting ensembles serve as null models to study the expected level of order given the fixed constraints. However, real systems are affected by a very large number of constraints and because they are so many, a list stating all of them one by one is unfeasible. Instead, we focus on their combined effect, which we extract by applying the generalized hypergeometric ensemble, gHypEG [15], to a given network representation. This allows to encode observed interaction preferences among every pair of individuals as biases in the edge formation. By this, we capture in the network ensemble the combined restriction of *all* constraints that manifest as interaction preferences between individuals.

Specific structural constraints can be measured for example by network measures such as modularity or nestedness. What we propose is a measure for the combined effect of such constraints in order to capture the potentiality of the analyzed organization. Clearly with our measure we can not consider the origin of individual constraints. However, our measure provides a description of the whole organization and how constrained it is overall.

The paper is organized as follows. In Section 2 we derive our measure of potentiality for a social organization based on the Shannon entropy of a probability distribution. In Section 3 we first explain how this probability distribution can be calculated for a Generalized Hypergeometric Ensemble. We then also show how to obtain the Shannon entropy of this distribution by means of a limit-case approximation, because direct computations are infeasible because of the large size of the ensemble. In Section 4 we measure the potentiality of 3 empirical social organizations and then compare the computed values across the organizations. Finally, in Section 5 we summarize our approach and comment on its strengths and weaknesses.

2. Quantifying the Potentiality of a Social Organization

2.1. Network Representation of a Social Organization

We adopt a network perspective to study social organizations. The nodes of the network represent individuals, and the observed interactions between them are represented by edges. If multiple interactions between the same pair of individuals occur, we consider them as separate edges, so-called *multi-edges* [16]. For simplicity, we will always refer to them as edges. In this article, we will focus on undirected edges without self-loops, however, the methodology discussed can easily be extended to encompass directed edges, and self-loops.

According to this perspective, the observation of a social organization composed of n individuals yields an network $\hat{g}$, with n nodes and m edges, where m is the number of observed interactions. The *state* of the social organization, instead, is defined by a network ensemble composed of all possible networks $S = \{g_0, \ldots, g_N\}$, that encompass all possible configurations the social organization could attain, with $\hat{g} \in S$.

As an example, Figure 1 illustrates every possible network for 3 nodes and an increasing number of edges m. While for 3 nodes and 2 edges there are 6 possible networks, for 3 edges already 10 networks result. For 10 nodes and 10 edges there would be more than $2 \cdot 10^{10}$ possible networks. The general expression for the number of possible networks is

$$\binom{\frac{n(n-1)}{2} + m - 1}{m} \tag{1}$$

where $n(n-1)/2$ denotes the number of combinations between n nodes. Equation (1) can be derived directly from the known formula for drawing unordered samples with replacement. The replacement is important because we consider multi-edges.

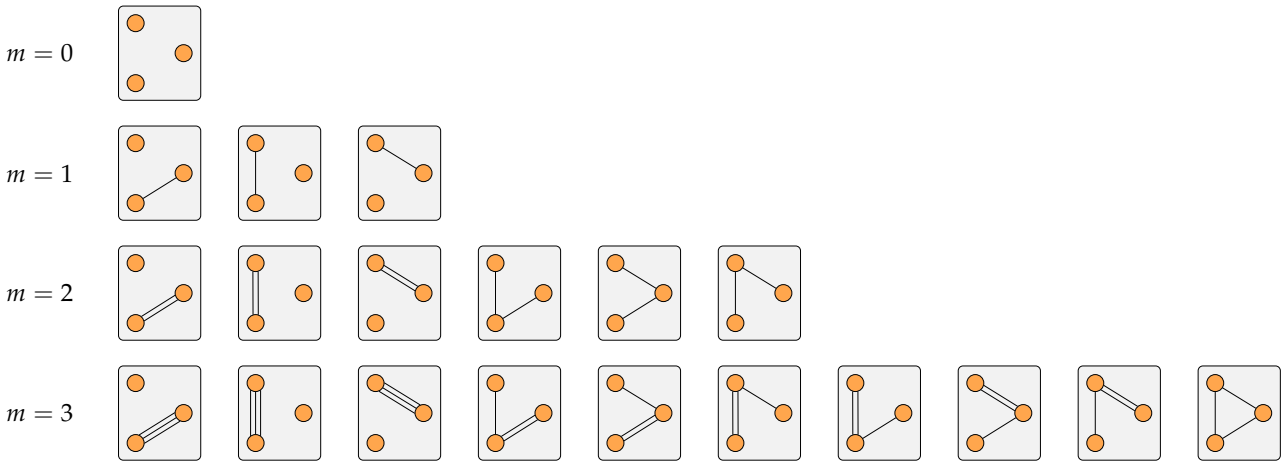

Figure 1. Visualization of the possible networks for 3 nodes and different numbers of edges. The edges are undirected and self-loops are not considered.

Notwithstanding the large number of possible networks, not all of them appear with the same probability. We denote with g a particular network, and by $P(g)$ the probability to find g, given n and m. A proper expression for $P(g)$ has to take into account that the ensemble, in addition to a fixed number of nodes and edges, also may have other constraints that need to be reflected in the probability distribution. This issue will be further discussed in Section 3. But, assuming that we have such an expression for $P(g)$, the information from this can be compressed by calculating Shannon entropy [17]

$$H = -\sum_{g \in S} P(g) \log P(g) \tag{2}$$

where S denotes the set of all possible networks for fixed n, m.

2.2. Potentiality of a Social Organization

Potentiality and constraints.

In our network representation, a large number of possible networks translates into a large number of possible configurations that can be attained by the social organization. Hence, we can use entropy to characterize the *potentiality* of the social organization, that is, its ability to attain these *different configurations* under the existing constraints. These constraints limit the number of configurations, i.e they reflect that a social organization cannot change from a given configuration to any arbitrary other configuration. Thus, *constraints lower the potentiality* of social organizations.

Such constraints can be temporal, i.e., they impose an order of occurrence to the edges in the network, as extensively examined in [8,18]. Or there can be spatial constraints that restrict the individuals in the choice of communication partners [19,20]. Social organizations can also be subject to hierarchical constraints [21], restricting e.g. the flow of information, or to social constraints [22] as discussed in Section 4.

How to proxy constraints.

We consider distributions $P(g)$ that capture communication biases among the individuals. These biases, or preferences, are the consequences of the constraints that restrict the social organization. We

take the observed number of interactions between each pair of individuals in a defined time interval as the proxy for the constraints. For this reason, we set the expected number of interactions between each pair of nodes in the ensemble to the observed ones. This choice ensures that the distribution $P(g)$ encodes the constraints in the ensemble, because we assume that constraints are expressed in the number of interactions between the nodes.

In the next Section we will demonstrate how to specify the probability distribution $P(g)$ characterizing the network ensemble such that this is achieved. To do so, we will employ the generalized hypergeometric ensemble (gHypEG) developed by Casiraghi and Nanumyan [15].

Network ensembles and their probability distribution.

What have we obtained by calculating Shannon entropy, i.e. a *single number* to characterize $P(g)$? To fully understand this, we have to recapture what information the probability distribution actually contains. $P(g)$ in fact characterizes the *diversity* of potential networks, i.e. the possible network configurations that can appear under the given constraints encoded in $P(g)$. We denote the totality of these configurations as the *network ensemble*. If there are only a *few* network configurations possible, the ensemble is comparably *small* and the resulting entropy is *low*. On the other hand, if many network configurations are possible, the ensemble becomes very large and the entropy is high.

3. Introducing the Generalized Hypergeometric Ensembles

3.1. Obtaining $P(g)$

For the calculation of Shannon entropy, Equation (2), we implicitly assumed that $P(g)$ is known. There are mainly two candidates for $P(g)$ that fit our requirements. One is the family of *exponential random graphs*, also known as ERGMs [23,24]. ERGMs follow an exponential distribution, thus it is possible to compute their Shannon entropy. Moreover they can incorporate a broad set of properties and constraints [25] which can fit virtually any characteristics of observed networks. However, ERGM fitting algorithms, especially when fitted to multi-edge networks, tend to not converge and thus cannot be efficiently computed for large networks.

Additionally, they are intended to consider a *predefined* set of constraints. But predefining all constraints of a social organization is unfeasible, given the very large number of constraints. Existing applications of ERGMs therefore examine specific constraints, as for example by fixing a clustering coefficient or degree assortativity [26]. However, we intend to measure the combined effect of the constraints, and therefore our choice is the second candidate, which is the *generalized hypergeometric ensemble* of random graphs [15,27] (gHypEG). This ensemble extends the configuration model (CM) [28] by encoding complex topological patterns, while at the same time preserving expected degree sequences.

Specifically, gHypEG keeps the number of nodes and edges fixed. However, different from the CM, the probability to connect two nodes depends not only on their (out- and in-) degrees (i.e., number of stubs), but also on an independent *propensity* of the two nodes to be connected, which captures *non-degree related effects* as explained in the following.

Parameters of a gHypEG.

The distribution of networks in a gHypEG is formulated in terms of two sets of parameters. The first set of parameters is represented in terms of the combinatorial matrix $\mathbf{\Xi}$ that encodes the CM. That means the entries Ξ_{ij} reflect all ways in which nodes i and j can be linked. As will be explained later in an undirected network without self-loops this number is $2\tilde{d}_i\tilde{d}_j$ for rescaled degrees $\tilde{d}_i$, $\tilde{d}_j$ of nodes i, j.

The second set of parameters is represented in terms of the propensity matrix $\mathbf{\Omega}$ which encodes preferences of nodes to be connected. That means, propensities allow to constrain the configuration model such that given edges are more likely than others, independently of the degrees of the respective

nodes. This creates a bias which is expressed by the ratio between any two elements Ω_{ij} and Ω_{kl}, i.e., the odds-ratio of observing an edge between nodes i and j instead of between k and l.

The matrices $\boldsymbol{\Xi}$ and $\boldsymbol{\Omega}$ both have dimension $n \times n$, where n is the number of nodes. The probability distribution that reflects the biased edge allocation described above is given by the multivariate Wallenius non-central hypergeometric distribution [29]. I.e., the probability of a network g in the gHypEG with parameters $\boldsymbol{\Xi}$ and $\boldsymbol{\Omega}$ is given as follows:

$$P(g|\boldsymbol{\Xi}, \boldsymbol{\Omega}) = \left[\prod_{i,j \in V, i<j} \binom{\Xi_{ij}}{A_{ij}} \right] \int_0^1 \prod_{i,j \in V, i<j} \left(1 - z^{\frac{\Omega_{ij}}{S_{\boldsymbol{\Omega}}}} \right)^{A_{ij}} dz \tag{3}$$

with

$$S_{\boldsymbol{\Omega}} = \sum_{i,j \in V, i<j} \Omega_{ij}(\Xi_{ij} - A_{ij}). \tag{4}$$

Equation (3) and Equation (4) hold for undirected networks without self-loops ($i < j$).

Calculating Ξ for networks.

We obtain the Ξ matrix for a given network according to Definition 4 and Lemma 3 in [15]. But, since in our applications there are no self-loops, we implement additional correction factors to preserve the expected degrees in the ensemble. Specifically, we ensure that the expected degrees are equal to the degrees in the initial network. The details can be found in Appendix A. Our Ξ_{ij} are therefore

$$\Xi_{ij} := \begin{cases} 2(d_i\theta_i)(d_j\theta_j) & \text{if } i < j \\ 0 & \text{else} \end{cases} \tag{5}$$

where d_i, d_j denote the degree of nodes i and j and the θ_i, θ_j denote the correction factors that ensure the expected degrees are preserved. In this definition the diagonal elements are 0 because we do not allow for self-loops. Also the entries in the lower triangular part are 0 to account for the networks being undirected.

Calculating Ω for networks.

We obtain the respective Ω matrix *for a given* Ξ *matrix* according to Corollary 7.3 in [15]. Thereby we ensure that, in addition to the expected degrees, even the expected numbers of edges between *all* pairs of nodes in the ensemble are equal to the respective numbers of edges in the initial network. Hence, our Ω_{ij} are

$$\Omega_{ij} := \begin{cases} \frac{1}{c} \log\left(1 - \frac{A_{ij}}{\Xi_{ij}}\right) & \text{if } i < j \\ 0 & \text{else} \end{cases} \tag{6}$$

where A_{ij} is the number of edges between nodes i and j, and c is a multiplicative constant which we choose such that the values in Ω are between 0 and 1 for simplicity. We refer to [15] for how special cases such as $A_{ij} = \Xi_{ij}$ can be handled. Again, the entries on the diagonal and in the lower triangular part of Ω are 0 to account for the networks having no self-loops and being undirected.

3.2. Tractability of the Entropy

Multinomial entropy approximation.

To compute the Shannon entropy of the fitted gHypEG according to Equations (2), (3), (4) is not straight-forward because of the very large number of networks in this ensemble. If we were to simply plug the probabilities of all networks into Equation (2), the very large number of summands quickly

becomes infeasible. Thus, instead of literally computing the entropy for a fitted gHypEG, we compute H using the fact that, for large networks, the multinomial distribution approximates the Wallenius distribution. The details of the derivation can be found in Appendix B. Hence, the gHypEG entropy can be approximated as

$$H^{\text{mult}} = -\log(m!) - m \sum_{i,j \in V, i<j} p_{ij} \log(p_{ij}) + \sum_{x=2}^{m} \sum_{i,j \in V, i<j} \binom{m}{x} p_{ij}^{x} (1-p_{ij})^{m-x} \log(x!) \tag{7}$$

where m is the number of edges in the network, V is the set of nodes, and

$$p_{ij} = \frac{\Xi_{ij}\Omega_{ij}}{\sum_{kl} \Xi_{kl}\Omega_{kl}} \tag{8}$$

Computing the multinomial entropy.

Equation (7) can be computed efficiently even for large ensembles. In SciPy [30] there exists an efficient implementation for computing the entropy of a given multinomial distribution. Our contribution is to apply this to approximate the entropy for a given gHypEG defined by Equations (7), (8).

3.3. Comparing Entropy Values

Normalizing value ranges.

The value range of Equation (7) depends on the number of nodes n and edges m. In particular, it is a known fact that Shannon entropy attains its maximum value $H^{\max}$ at equiprobability [17]. Hence, the entropy values are always in the interval $[0, H^{\max}]$.

For undirected networks without self-loops equiprobability corresponds to

$$p_{ij}^{\max} = \frac{2}{n(n-1)} \tag{9}$$

i.e., all possible pairs of nodes can be chosen with the same probability. For two different ensembles, however, $H^{\max}$ can be different because it depends on n and on m via Equation (7). To compare the values of H^{mult}, Equation (7), we normalize them by their respective maximum values:

$$H^{\text{norm}} := \frac{H^{\text{mult}}}{H^{\max}} \equiv \hat{H} \in [0,1] \tag{10}$$

A small value means that the ensemble contains only very few networks, given the constraints. With respect to the p_{ij} this means that only very few have probabilities considerably different from zero. A large value, on the other hand, means that pairs of nodes are chosen almost at random, because of the very few constraints. Hence, $\hat{H}$ indeed reflects the potentiality of the social organization, namely its ability to attain *different configurations* under given constraints.

3.4. Examples for $\hat{H}$

Two special cases.

To illustrate how constraints can be encoded in the ensemble, we use two examples, a complete network and a star network (see Figure 2) for which we consider undirected edges and no self-loops. We fit the Ξ and Ω matrices according to Equations (5), (6).

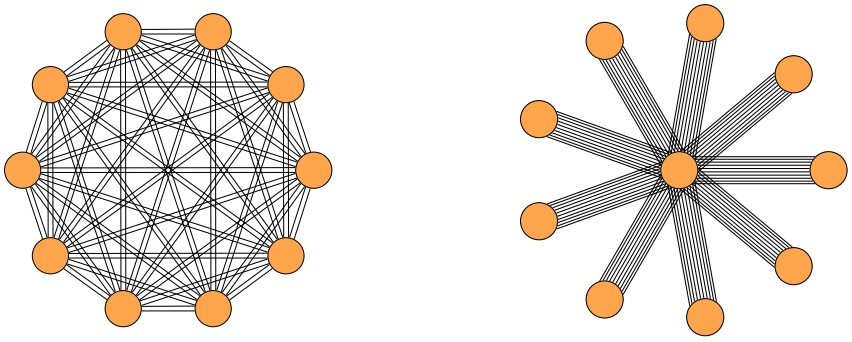

Figure 2. The two networks from Section 3.4. Left: Multi-edge complete network. Right: Multi-edge star network.

Complete network.

This network has 10 nodes and 90 edges. Because we consider a multi-edge network, each node has 2 edges to every other node. This results in $d_i = d_j = 18$. The correction factors are obtained by solving the set of equations in the Appendix A and yield $\theta_i = \theta_j = 1.054$. The resulting network is depicted in Figure 2 (left), and the resulting Ξ and Ω matrices are stated in full in Appendix C. The entries of the Ξ matrix for this network are computed according to Equation (5) as

$$\Xi^{\mathrm{C}}_{ij} = \begin{cases} 720 & \text{if } i < j \\ 0 & \text{else} \end{cases} \tag{11}$$

Since in the complete network every node has the same number of edges to every other node, there are no preferences for specific pairs of nodes. Therefore one way to choose the Ω matrix to encode *no bias* is

$$\Omega^{\mathrm{C}}_{ij} = \begin{cases} 1 & \text{if } i < j \\ 0 & \text{else} \end{cases} \tag{12}$$

which corresponds to Theorem 8 in [15] for an undirected network without self-loops. Remember that Ω_{ij}/Ω_{kl} is the odds-ratio of observing an edge between nodes i and j instead of nodes k and l. By choosing Ω_{ij} according to Equation (12) such ratios are always equal to 1. By plugging Ξ^{C} and Ω^{C} into Equation (10) we obtain $\hat{H} = 1$. This means that there are no edge-preferences between particular pairs of nodes, which is trivial because the example was chosen as such.

Star network.

This network has again 10 nodes and 90 edges. But this time there is 1 center node and 9 peripheral nodes, i.e., the network has the constraint that each peripheral node has 10 edges that are all attached to the center node, as depicted in Figure 2 (right). This results in a degree $d_i = 90$ for the center node placed at $i = 1$ and degrees $d_j = 10$ for all peripheral nodes $j \neq 1$. Again, the Ξ and Ω matrices for this network are stated in full in Appendix C.

When computing the Ξ matrix according to Equation (5) we obtain

$$\Xi^{\mathrm{S}}_{ij} = \begin{cases} 3592 & \text{if } i = 1, j > 1 \\ 2 & \text{if } i > 1, j > i \\ 0 & \text{else} \end{cases} \tag{13}$$

where the center node is placed at $i = 1$.

The other matrix, Ω, has to reproduce the constraint that peripheral nodes can only communicate with the center node. One choice to fulfil this is

$$\Omega^{\mathrm{S}}_{ij} = \begin{cases} 1 & \text{if } i = 1 \text{ and } i < j \\ 0 & \text{else} \end{cases} \tag{14}$$

This choice of Ω specifies that observing an edge from node 1 (the center) to any two peripheral nodes k or l occurs with the same probability, because the odds-ratio Ω_{1k}/Ω_{1l} is equal to 1. On the other hand, the odds-ratio for a peripheral node i to form a link with another peripheral node k instead of with the center node 1, namely Ω_{ik}/Ω_{1i}, is 0 (or infinity if the inverse ratio is formed). This encodes the constraint that all edges have to be incident to the center node.

By plugging Ξ^{S} and Ω^{S} into Equation (10) we obtain $\hat{H} = 0.27$ for our star network. This value is much lower than for the complete network and reflects the very restrictive constraint that all edges have to be incident to the center node.

4. Applications to Real-World Data Sets

4.1. Examined Data Sets

In this Section we apply our potentiality measure to 5 empirical networks of social organizations. These networks were constructed from publicly available data sets which we shortly describe in the following.

Southern Women data set.

It was introduced by Davis *et al.* [31] and contains information about 18 women and their participation in 14 social events. Instead of constructing a bipartite network, we use a so-called one-mode representation (i.e. a specific projection of the bipartite network) in which the women correspond to the nodes and the edges correspond to co-participations in the social events. There are no self-loops in this network and edges are undirected.

Karate Club data set.

It was introduced by Zachary [2] and the network contains 34 nodes corresponding to the members of this university Karate club. Edges correspond to co-participation of members in different activities. They are all undirected and there are no self-loops. There are 8 activities considered, thus the number of possible edges between any pair of nodes is less or equal than 8. In total, there are 231 edges.

Conference data set.

This data set is part of the SocioPatterns project. It contains data about interactions among conference participants during the ACM Hypertext 2009 conference. [3] To measure the interactions the participants were wearing proximity sensors. For each interaction between two participants the measured information contains their anonymous ids as well as the time of the respective measurement. From this information we constructed 3 networks, one for each day of the conference. In each network the nodes correspond to the 113 participants in the data and the edges correspond to their interactions at the respective day. None of the networks contains self-loops and all edges are undirected. All 3 networks have the same set of nodes but differ slightly in the number of edges as can be seen in Table 1.

Network overview.

To summarize the networks Table 1 lists the general network statistics besides the computed potentiality values of $\hat{H}$. Furthermore, all networks are visualized in Figure 3. This Figure already suggests that the networks are structurally different. For example, the Karate Club network shows a cluster structure which is not apparent in the Southern Women network. And all Conference networks have isolated nodes which neither the Karate Club network nor the Southern Women network have.

Table 1. Network statistics of the 5 examined empirical networks. n and m denote the number of nodes and edges in each network. m/n is the average number of multi-edges per node. D is the density of the network, i.e. the number of linked node pairs normalized to the total number of possible node pairs, after reducing all multi-edges into single edges. $\hat{H}$ denotes the normalized entropy computed according to Equation (10). $\hat{H}_{\text{gcc}}$ corresponds to $\hat{H}$ when only the largest connected component in each network is considered. All networks are undirected and have no self-loops.

Network	n	m	m/n	D	$\hat{H}$	$\hat{H}_{\text{gcc}}$
Southern Women	18	322	17.89	0.91	**0.89**	**0.89**
Karate Club	34	231	6.79	0.14	**0.31**	**0.31**
Conference $t = 1$	113	6925	61.28	0.15	**0.21**	**0.24**
Conference $t = 2$	113	7131	63.11	0.17	**0.22**	**0.25**
Conference $t = 3$	113	6762	59.84	0.15	**0.19**	**0.23**

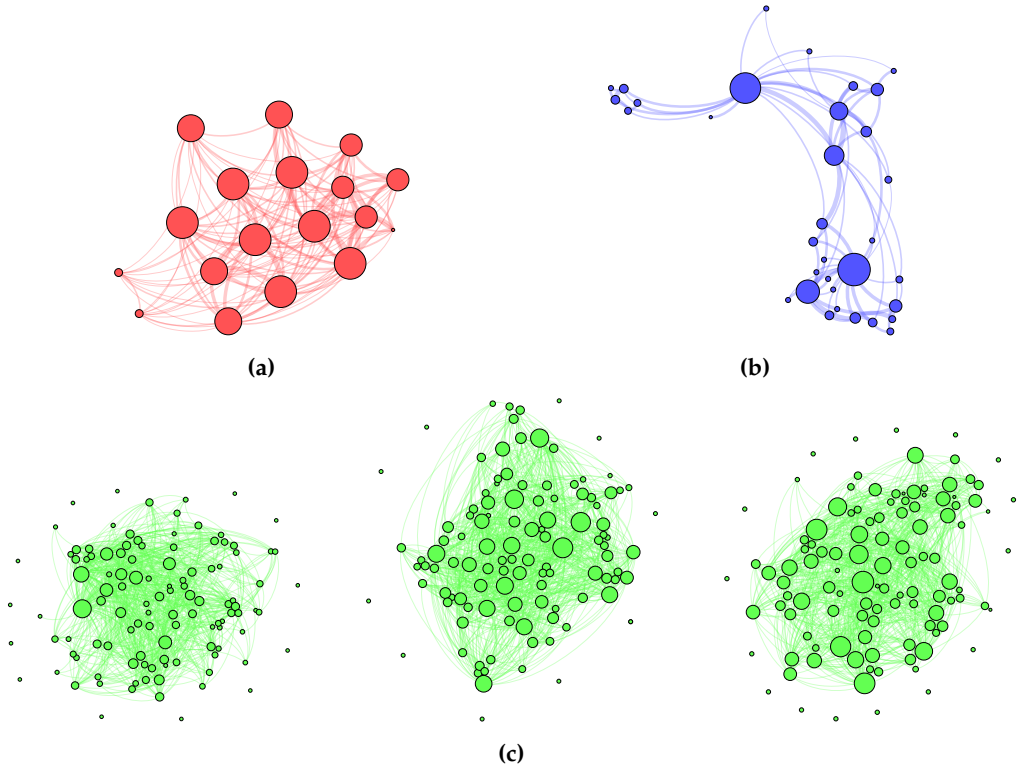

Figure 3. Network visualizations: (a) Southern Women (red), (b) Karate Club (blue), (c) Conference (green) at day 1 (left), day 2 (center), and day 3 (right). The node size is proportional to the degree.

4.2. Potentiality of the Empirical Networks

For each of the 5 empirical networks we computed the potentiality $\hat{H}$ as outlined in Sections 2 and 3. The computed values for $\hat{H}$ are listed in Table 1. In the following we comment on the results.

Southern Women network.

This network attains a very high potentiality at around $\hat{H} = 0.9$, meaning that there are only few constraints in the women's interaction. In fact, there are almost no preferences for specific pairs of women. Instead, everyone interacts with everyone else in a rather homogeneous way. The absence of a preference structure in co-attending events is also visible in the network plot in Figure 3, which looks similar to the complete network considered in Section 3.4. This corresponds to the high density, D, in Table 1. Note also that the measured potentiality is high, but it is still not at its maximum value 1, i.e. there are constraints present in the network that restrict interactions. Evidence for these constraints are the two groups that were identified among the southern women in [31,32].

Karate Club network.

This network results in a lower potentiality at around $\hat{H} = 0.3$ which indicates that the network is more restricted by constraints. Indeed, it is known that two social groups, which both had their own leader, co-existed in the Karate club. Most interactions among the club members occurred within the groups, and in particular with the respective group leaders. These restrictions explain why the potentiality of this network is not particularly large, especially when compared to the Southern Women network.

Conference networks.

The lowest potentialities are attained at around 0.2 by the networks of the conference participants. For each network we observed that the nodes had high degrees, because of the multi-edges, but were linked only to a *few* other nodes (i.e. a rather sparse network). This implies a relatively strong preferential linkage between *specific* pairs of nodes. On the other hand, on all 3 days of the conference, there was at least one individual who communicated with at least 50% of conference participants in the data set (probably the conference organizer). This induces a star-like interaction effect. But, there were also a smaller number of isolated nodes which were, on a given day, not involved in any interaction.

Isolated nodes decrease the potentiality because in the theoretical maximum entropy $H^{\max}$ they have to be considered. When omitting these isolated nodes, we still find only slightly higher values of the potentialities around 0.24 because of their small number. Furthermore, it is remarkable that the networks of all three days have similar potentialities. One could have expected that on day 3 of the conference, participants mainly interact with those they already know. But this is obviously not the case.

5. Conclusions

In this paper we address the question how many different states a social organization can attain. Why is this of importance? We argue that the number of such possible states is an indication of the ability of the organization to respond to various influences. As there can be a vast variety of such influences, the corresponding number ideally should be very large. This indicates that, even for unforeseeable events, the social organization still has many ways to respond. We call such an ability the *potentiality* of the organization.

To quantify this potentiality, we need an appropriate representation of the social organization. In this paper, we choose a network approach, where nodes represent individuals and edges their repeated interactions. This leads to a multi-edge network. A network ensemble then contains all possible networks that fulfill a given set of constraints. Such constraints are detected from the observed network and encoded as *propensities*, i.e. as interaction preferences. The statistical ensemble of all possible networks is then given by the generalized hypergeometric ensemble (gHypEG). From this, we can calculate a Shannon entropy, which is used to proxy the potentiality of the organization.

In the following, we comment further on the strengths and weaknesses of our approach.

Fixed numbers of nodes and edges.

We focus on ensembles with a fixed number of nodes and edges, hence imposing that only networks of this size are attainable by the organization. Thereby we neglect system growth on purpose, to provide a general measure of potentiality.

Large number of degrees of freedom.

Using gHypEG, we are able to consider the maximum possible degrees of freedom, meaning that every detail is modeled in the Ξ and Ω matrices. This way, we obtain a high model complexity. A more refined approach could be to compare ensembles of various complexities based on goodness-of-fit measures such as AIC or BIC. Thereby also simpler ensembles could be involved that, for example,

consider communication preferences only between certain *communities* in the network. Such choices of simpler ensembles were not considered because, again, we want to provide a general approach not restricted to systems with particular community structures.

Computability.

For social organizations with only 10 individuals and 10 interactions there are already more than $2 \cdot 10^{10}$ possible network representations in the ensemble. According to Equation (2), all of these networks have to be considered individually to compute the Shannon entropy. Hence, even simple approaches to directly compute the entropy are computationally infeasible already for very small organizations. Our approach instead uses that the Wallenius distribution underlying the gHypEG converges to a multinomial distribution in the limit of large networks for which the entropy can be computed efficiently. This allows to study the potentiality of a wide range of social organizations.

Our main methodological contribution is indeed the novel way to conceptualize potentiality for a social organization using its representation as a multi-edge network. As long as this representation is justified, our approach can be extended to other systems.

Author Contributions: All authors participated in the writing of the paper. The code for the analysis was written by Christian Zingg.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A Fitting the Ξ Matrix for Undirected Networks Without Self-Loops

In the case of undirected networks, the Ξ matrix can be derived according to Definition 4 and Lemma 3 in [15], i.e. as $2d_i d_j$ for the degrees d_i and d_j of nodes i and j. If we further disallow self-loops, an additional step is necessary to ensure that the expected degrees in the ensemble still correspond to those of the initial network, and the probability spaces are comparable. In particular, we need to ensure that the entries of the Ξ matrix sum to m^2 in the case of directed networks, and to $4m^2$ in the case of undirected networks.

To do so, in the case of undirected networks we define the entries of Ξ as follows:

$$\Xi_{ij} := 2d_i d_j \theta_i \theta_j, \tag{A1}$$

where θ_i is the correction factor corresponding to node i. To estimate the parameters θ_i we fix the two constraints just described: (i) degrees have to be preserved in expectation, (ii) entries of Ξ sum to $4m^2$. This gives the following system of equations:

$$\begin{cases} \sum_{i<j} \Xi_{ij} = 4m^2 & \\ \frac{m}{\sum_{l<k} \Xi_{lk}} \sum_{j \neq 1} \Xi_{1j} & = d_1 \\ & \vdots \\ \frac{m}{\sum_{l<k} \Xi_{lk}} \sum_{j \neq n} \Xi_{nj} & = d_n \end{cases} \tag{A2}$$

where $m \sum_j \Xi_{ij} / \sum_{l<k} \Xi_{lk}$ gives the expected degree of node i according to the configuration model [15]. By inputting Equation A2 in the system above, we can simplify it into the following system of n equations in n variables for which we find a numerical solution.

$$\begin{cases} 2m\theta_1 \sum_{j \neq 1} d_j \theta_j & = 0 \\ & \vdots \\ 2m\theta_n \sum_{j \neq n} d_j \theta_j & = 0 \end{cases} \tag{A3}$$

For the case of the *star network* discussed in Section 3.4, there is no exact solution to this set of equations. It is not possible to fix the number of edges to 90 while simultaneously fixing the expected degrees over a gHypEG ensemble to the observed degrees in the star network. Therefore, we introduce an approximation that allows for a small error tolerance between a degree d_i and its corresponding expected degree $\tilde{d}_i$ in the ensemble:

$$\left|d_i - \tilde{d}_i\right| \leq 0.5 \tag{A4}$$

This allows us to obtain a solution close enough to fulfill the conditions of the equation system.

Appendix B Convergence in Distribution of gHypEGs

In the following discussion, we show a short theorem that provides the limiting distribution for gHypEGs. Recall that gHypEGs are described by the sampling **without replacement** of m edges from an urn containing $M = \sum_{ij} \Xi_{ij} = m^2$ edges. When M is large, and some other constraints are met, the gHypEGs sampling process can be approximated by a sampling **with replacement**. Here, we provide a rigorous demonstration of this statement. Note that it is a known result that the the hypergeometric distribution (i.e., an urn sampling without replacement) converges to the multinomial distribution (i.e., an urn sampling with replacement). However, to our best knowledge, the are no analytic proofs that the Wallenius non-central hypergeometric distribution also converges to the multinomial distribution. Hence, we now prove that the sampling with competition described by Wallenius' multivariate non-central hypergeometric distribution can be approximated by a multinomial distribution with probabilities defined as in Equation (8).

Theorem A1 (Convergence of Wallenius' distribution to the multinomial distribution)**.** *Let X be a random variable distributed according to Wallenius' multivariate hypergeometric non-central distribution with parameters $\boldsymbol{\Xi}$, $\boldsymbol{\Omega}$, and m, given by*

$$P(X = \mathbf{A}) = \left[\prod_{i,j \in V} \binom{\Xi_{ij}}{A_{ij}}\right] \int_0^1 \prod_{i,j \in V} \left(1 - t^{\frac{\Omega_{ij}}{S_{\boldsymbol{\Omega}}}}\right)^{A_{ij}} dt \tag{A5}$$

with

$$S_{\boldsymbol{\Omega}} = \sum_{l,k \in V} \Omega_{lk}(\Xi_{lk} - A_{lk}). \tag{A6}$$

Let $\Xi_{ij} = n\tilde{\Xi}_{ij}$ $\forall i, j \in V$ such that

$$\frac{\Omega_{ij}\Xi_{ij}}{\sum_{(l,k) \in V \times V} \Omega_{lk}\Xi_{lk}} = \frac{\Omega_{ij}\tilde{\Xi}_{ij}}{\sum_{(l,k) \in V \times V} \Omega_{lk}\tilde{\Xi}_{lk}} = p_{ij} \quad \forall i, j \in V. \tag{A7}$$

Then, the Wallenius' multivariate non-central hypergeometric distribution converges to the multinomial distribution with probabilities p_{ij}:

$$P(X = \mathbf{A}) \to \frac{m!}{\prod_{(i,j) \in V \times V} A_{ij}!} \prod_{(i,j) \in V \times V} (p_{ij})^{A_{ij}} \quad \text{as } n \to \infty \tag{A8}$$

Proof. From now on, we will write $\prod_{(i,j) \in V \times V}$ as $\prod_{i,j \in V}$ and $\sum_{(i,j) \in V \times V}$ as $\sum_{i,j \in V}$. With this notation the Wallenius' multivariate non-central hypergeometric distribution is

$$P(X=\mathbf{A}) = \left[\prod_{i,j\in V}\binom{\Xi_{ij}}{A_{ij}}\right]\int_0^1 \prod_{i,j\in V}\left(1-t^{\frac{\Omega_{ij}}{S_\Omega}}\right)^{A_{ij}} dt \tag{A9}$$

$$= \frac{m!}{M^m}\left[\prod_{i,j\in V}\binom{\Xi_{ij}}{A_{ij}}\right]\cdot\frac{M^m}{m!}\int_0^1 \prod_{i,j\in V}\left(1-t^{\frac{\Omega_{ij}}{S_\Omega}}\right)^{A_{ij}} dt \tag{A10}$$

where we obtain Equation (A10) by multiplying and dividing by M^m and $m!$. The first term of Equation (A10) can be re-written in the following form:

$$\frac{m!}{M^m}\prod_{i,j\in V}\binom{\Xi_{ij}}{A_{ij}} = \frac{m!}{M^m}\prod_{i,j\in V}\frac{\Xi_{ij}!}{A_{ij}!\,(\Xi_{ij}-A_{ij})!} \tag{A11}$$

$$= \frac{m!}{\prod_{i,j\in V}A_{ij}!}\prod_{i,j\in V}\frac{\Xi_{ij}!}{M^{A_{ij}}\,(\Xi_{ij}-A_{ij})!} \tag{A12}$$

$$= \frac{m!}{\prod_{i,j\in V}A_{ij}!}\prod_{i,j\in V}\left(\prod_{k=1}^{A_{ij}}\frac{\Xi_{ij}-A_{ij}+k}{M}\right) \tag{A13}$$

Note that to obtain Equation (A12) we have written $M^m = \prod_{i,j\in V} M^{A_{ij}}$ which follows from the fact that $\sum_{i,j\in V} A_{ij} = m$. Let now $\Xi_{ij} = n\tilde{\Xi}_{ij}$ and $\tilde{M} = \sum_{i,j\in V}\tilde{\Xi}_{ij}$ such that $M = \sum_{i,j\in V}\Xi_{ij} = n\sum_{i,j\in V}\tilde{\Xi}_{ij} = n\tilde{M}$ and $\tilde{\Xi}_{ij}/\tilde{M} = \Xi_{ij}/M = \tilde{p}_{ij}$. We substitute $\tilde{\Xi}_{ij}$ and $\tilde{M}$ in Equation (A13) and then, we can calculate its limit for $n\to\infty$:

$$\lim_{n\to\infty}\frac{m!}{\prod_{i,j\in V}A_{ij}!}\prod_{i,j\in V}\left(\prod_{k=1}^{A_{ij}}\frac{n\tilde{\Xi}_{ij}-A_{ij}+k}{n\tilde{M}}\right) = \frac{m!}{\prod_{i,j\in V}A_{ij}!}\prod_{i,j\in V}\left(\prod_{k=1}^{A_{ij}}\lim_{n\to\infty}\frac{n\tilde{\Xi}_{ij}-A_{ij}+k}{n\tilde{M}}\right) \tag{A14}$$

$$= \frac{m!}{\prod_{i,j\in V}A_{ij}!}\prod_{i,j\in V}\tilde{p}_{ij}^{A_{ij}} \tag{A15}$$

We are left with second term of Equation (A10) that contains the integral. To evaluate this term, we substitute $\Xi_{ij} = n\tilde{\Xi}_{ij}$ and $M = n\tilde{M}$ and calculate its limit for $n\to\infty$:

$$\lim_{n\to\infty}\frac{M^m}{m!}\int_0^1\prod_{i,j\in V}\left(1-t^{\frac{\Omega_{ij}}{\sum\Omega}}\right)^{A_{ij}}dt=\lim_{n\to\infty}\frac{(n\tilde{M})^m}{m!}\int_0^1\prod_{i,j\in V}\left(1-t^{\frac{\Omega_{ij}}{\sum_{lk}\Omega_{lk}(n\tilde{\Xi}_{lk}-A_{lk})}}\right)^{A_{ij}}dt \tag{A16}$$

$$=\frac{\tilde{M}^m}{m!}\cdot\int_0^1\lim_{n\to\infty}n^m\prod_{i,j\in V}\left(1-t^{\frac{\Omega_{ij}}{\sum_{lk}\Omega_{lk}(n\tilde{\Xi}_{lk}-A_{lk})}}\right)^{A_{ij}}dt \tag{A17}$$

$$=\frac{\tilde{M}^m}{m!}\cdot\int_0^1\prod_{i,j\in V}\lim_{n\to\infty}n^{A_{ij}}\left(1-t^{\frac{\Omega_{ij}}{\sum_{lk}\Omega_{lk}(n\tilde{\Xi}_{lk}-A_{lk})}}\right)^{A_{ij}}dt \tag{A18}$$

$$=\frac{\tilde{M}^m}{m!}\cdot\int_0^1\prod_{i,j\in V}\left(\lim_{n\to\infty}n\left(1-t^{\frac{\Omega_{ij}}{\sum_{lk}\Omega_{lk}(n\tilde{\Xi}_{lk}-A_{lk})}}\right)\right)^{A_{ij}}dt \tag{A19}$$

$$=\frac{\tilde{M}^m}{m!}\cdot\int_0^1\prod_{i,j\in V}\left(-\frac{\Omega_{ij}\log(t)}{\sum_{lk}\tilde{\Xi}_{lk}\Omega_{lk}}\right)^{A_{ij}}dt \tag{A20}$$

$$=\prod_{i,j\in V}\left(\frac{\Omega_{ij}\tilde{M}}{\sum_{lk}\tilde{\Xi}_{lk}\Omega_{lk}}\right)^{A_{ij}}\cdot\frac{1}{m!}\int_0^1\left(\log\frac{1}{t}\right)^m dt \tag{A21}$$

$$=\prod_{i,j\in V}\left(\frac{\Omega_{ij}\tilde{M}}{\sum_{lk}\tilde{\Xi}_{lk}\Omega_{lk}}\right)^{A_{ij}} \tag{A22}$$

Note that Equation (A17) follows by the Lebesgue dominated convergence theorem and by the finiteness of the factors in the integral. To obtain Equation (A20) we have used l'Hôpital's rule, i.e. by recalling that $\lim_{n\to\infty}(1-t^{a/n})n=\lim_{x\to 0}(1-t^{ax})/x=\lim_{x\to 0}(-a\log(t)t^{ax})=-a\log(t)$. Equation (A22) follows from an integral definition of the Γ function, precisely we have used $\Gamma(z+1)=\int_0^1\log(1/t)^z dt=z!$. Finally, by joining Equation (A15) and Equation (A22), we obtain the limit of Equation (A9):

$$\frac{m!}{\prod_{i,j\in V}A_{ij}!}\prod_{i,j\in V}\tilde{p}_{ij}^{A_{ij}}\prod_{i,j\in V}\left(\frac{\Omega_{ij}\tilde{M}}{\sum_{lk}\tilde{\Xi}_{lk}\Omega_{lk}}\right)^{A_{ij}}=\frac{m!}{\prod_{i,j\in V}A_{ij}!}\prod_{i,j\in V}\left(\tilde{p}_{ij}\cdot\frac{\Omega_{ij}\tilde{M}}{\sum_{lk}\tilde{\Xi}_{lk}\Omega_{lk}}\right)^{A_{ij}} \tag{A23}$$

$$=\frac{m!}{\prod_{i,j\in V}A_{ij}!}\prod_{i,j\in V}\left(\frac{\Omega_{ij}\tilde{\Xi}_{ij}}{\sum_{lk}\tilde{\Xi}_{lk}\Omega_{lk}}\right)^{A_{ij}} \tag{A24}$$

$$=\frac{m!}{\prod_{i,j\in V}A_{ij}!}\prod_{i,j\in V}(p_{ij})^{A_{ij}} \tag{A25}$$

where Equation (A24) follows from the definition of $\tilde{p}_{ij}=\tilde{\Xi}_{ij}/\tilde{M}$ and Equation (A25) follows from the definition of p_{ij}. □

Appendix C Full Matrices for the Complete and the Star Network

This appendix contains the full Ξ and Ω matrices of the gHypEG fits for the two networks described in Section 3.4.

The matrices of the complete network are

$$\Xi^{\mathrm{C}} = \begin{pmatrix} 0 & 720 & 720 & 720 & \dots & 720 \\ 0 & 0 & 720 & 720 & & \\ 0 & 0 & 0 & 720 & & \vdots \\ 0 & 0 & 0 & 0 & \ddots & \\ \vdots & & & \ddots & \ddots & 720 \\ 0 & & \dots & & 0 & 0 \end{pmatrix} \qquad \Omega^{\mathrm{C}} = \begin{pmatrix} 0 & 1 & 1 & 1 & \dots & 1 \\ 0 & 0 & 1 & 1 & & \\ 0 & 0 & 0 & 1 & & \vdots \\ 0 & 0 & 0 & 0 & \ddots & \\ \vdots & & & \ddots & \ddots & 1 \\ 0 & & \dots & & 0 & 0 \end{pmatrix} \tag{A26}$$

The matrices of the star network are

$$\Xi^{\mathrm{S}} = \begin{pmatrix} 0 & 3592 & 3592 & 3592 & \dots & 3592 \\ 0 & 0 & 2 & 2 & \dots & 2 \\ 0 & 0 & 0 & 2 & & 2 \\ 0 & 0 & 0 & 0 & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 2 \\ 0 & & \dots & & 0 & 0 \end{pmatrix} \qquad \Omega^{\mathrm{S}} = \begin{pmatrix} 0 & 1 & 1 & 1 & \dots & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & & \\ 0 & 0 & 0 & 0 & & \vdots \\ \vdots & & & & \ddots & 0 \\ 0 & & \dots & & 0 & 0 \end{pmatrix} \tag{A27}$$

The Ξ^{S} is only an approximation as outlined in Appendix A. It is needed because the equation-system to determine the correction factors has no exact solution for the star network.

References

1. Schweitzer, F.; Nanumyan, V.; Tessone, C.J.; Xia, X. How do OSS projects change in number and size? A large-scale analysis to test a model of project growth. *Advances in Complex Systems* **2014**, *17*, 1550008. doi:10.1142/S0219525915500083.
2. Zachary, W.W. An Information Flow Model for Conflict and Fission in Small Groups. *Journal of Anthropological Research* **1977**, *33*, 452–473, [NIHMS150003]. doi:10.1086/jar.33.4.3629752.
3. Isella, L.; Stehlé, J.; Barrat, A.; Cattuto, C.; Pinton, J.F.; Van den Broeck, W. What's in a crowd? Analysis of face-to-face behavioral networks. *Journal of Theoretical Biology* **2011**, *271*, 166–180. doi:10.1016/j.jtbi.2010.11.033.
4. Dorogovtsev, S.; Mendes, J. *Evolution of Networks*; Vol. 57, Oxford University Press, 2003; [arXiv:cond-mat/0106144]. doi:10.1093/acprof:oso/9780198515906.001.0001.
5. Newman, M.E.J. The Structure and Function of Complex Networks. *SIAM Review* **2003**, *45*, 167–256. doi:10.1137/S003614450342480.
6. Ebeling, W.; Freund, J.; Schweitzer, F. *Komplexe Strukturen: Entropie und Information*; Vieweg+Teubner Verlag: Wiesbaden, 1998. doi:10.1007/978-3-322-85167-3.
7. Zhao, K.; Karsai, M.; Bianconi, G. Entropy of dynamical social networks. *PLoS ONE* **2011**, *6*, e28116, [1201.3292]. doi:10.1371/journal.pone.0028116.
8. Kulisiewicz, M.; Kazienko, P.; Szymanski, B.K.; Michalski, R. Entropy Measures of Human Communication Dynamics. *Scientific Reports* **2018**, *8*, 15697, [1801.04528]. doi:10.1038/s41598-018-32571-3.
9. Santiago-Paz, J.; Torres-Roman, D.; Velarde-Alvarado, P. Detecting anomalies in network traffic using Entropy and Mahalanobis distance. CONIELECOMP 2012, 22nd International Conference on Electrical Communications and Computers. IEEE, 2012, pp. 86–91. doi:10.1109/CONIELECOMP.2012.6189887.
10. Rajaram, R.; Castellani, B. An entropy based measure for comparing distributions of complexity. *Physica A* **2016**, *453*, 35–43. doi:10.1016/j.physa.2016.02.007.
11. Hanel, R.; Thurner, S. A comprehensive classification of complex statistical systems and an axiomatic derivation of their entropy and distribution functions. *Europhysics Letters* **2011**, *93*, 20006, [1005.0138]. doi:10.1209/0295-5075/93/20006.

12. Bianconi, G. Entropy of network ensembles. *Physical Review E* **2009**, *79*, 036114. doi:10.1103/PhysRevE.79.036114.
13. Coon, J.P.; Dettmann, C.P.; Georgiou, O. Entropy of spatial network ensembles. *Physical Review E* **2018**, *97*, 042319. doi:10.1103/PhysRevE.97.042319.
14. Bianconi, G. The entropy of randomized network ensembles. *Europhysics Letters* **2008**, *81*, 28005. doi:10.1209/0295-5075/81/28005.
15. Casiraghi, G.; Nanumyan, V. Generalised hypergeometric ensembles of random graphs: the configuration model as an urn problem. *arXiv preprint arXiv:1810.06495* **2018**.
16. Bollobás, B. *Modern Graph Theory*; Vol. 184, *Graduate Texts in Mathematics*, Springer New York: New York, NY, 1998. doi:10.1007/978-1-4612-0619-4.
17. Shannon, C.E. A Mathematical Theory of Communication. *Bell System Technical Journal* **1948**, *27*, 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.
18. Scholtes, I. When is a Network a Network? Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '17; ACM Press: New York, New York, USA, 2017; pp. 1037–1046. doi:10.1145/3097983.3098145.
19. Vaccario, G.; Verginer, L.; Schweitzer, F. Reproducing scientists' mobility: A data-driven model. *arXiv preprint arXiv:1811.07229* **2018**, [1811.07229].
20. Liben-Nowell, D.; Novak, J.; Kumar, R.; Raghavan, P.; Tomkins, A. Geographic routing in social networks. *Proceedings of the National Academy of Sciences* **2005**, *102*, 11623–11628. doi:10.1073/pnas.0503018102.
21. Zanetti, M.S.; Scholtes, I.; Tessone, C.J.; Schweitzer, F. The rise and fall of a central contributor: Dynamics of social organization and performance in the GENTOO community. 2013 6th International Workshop on Cooperative and Human Aspects of Software Engineering, CHASE 2013 - Proceedings. IEEE, 2013, pp. 49–56, [1302.7191]. doi:10.1109/CHASE.2013.6614731.
22. Scholtes, I.; Mavrodiev, P.; Schweitzer, F. From Aristotle to Ringelmann: a large-scale analysis of team productivity and coordination in Open Source Software projects. *Empirical Software Engineering* **2016**, *21*, 642–683. doi:10.1007/s10664-015-9406-4.
23. Krivitsky, P.N.; Butts, C.T. Exponential-family Random Graph Models for Rank-order Relational Data. *Sociological Methodology* **2017**, p. 008117501769262. doi:10.1177/0081175017692623.
24. Park, J.; Newman, M.E. Statistical mechanics of networks. *Physical Review E* **2004**, *70*, 13, [arXiv:cond-mat/0405566]. doi:10.1103/PhysRevE.70.066117.
25. Morris, M.; Handcock, M.S.; Hunter, D.R. Specification of Exponential-Family Random Graph Models: Terms and Computational Aspects. *Journal of Statistical Software* **2008**, *24*, [NIHMS150003]. doi:10.18637/jss.v024.i04.
26. Fischer, R.; Leitão, J.C.; Peixoto, T.P.; Altmann, E.G. Sampling Motif-Constrained Ensembles of Networks. *Physical Review Letters* **2015**, *115*, 188701. doi:10.1103/PhysRevLett.115.188701.
27. Casiraghi, G.; Nanumyan, V.; Scholtes, I.; Schweitzer, F. Generalized Hypergeometric Ensembles: Statistical Hypothesis Testing in Complex Networks. *arXiv preprint arXiv:1607.02441* **2016**, [1607.02441].
28. Molloy, M.; Reed, B. A critical point for random graphs with a given degree sequence. *Random Structures & Algorithms* **1995**, *6*, 161–180. doi:10.1002/rsa.3240060204.
29. Wallenius, K.T. Biased Sampling: the Noncentral Hypergeometric Probability Distribution. Ph.d. thesis, Stanford University, 1963.
30. Jones, E.; Oliphant, T.; Peterson, P.; others. SciPy: Open source scientific tools for Python, 2001–. [Online; accessed 2019-01-28].
31. Davis, A.; Gardner, B.B.; Gardner, M.R. *Deep South; a Social Anthropological Study of Caste and Class*; The University of Chicago Press: Chicago, 1941.
32. Doreian, P. On the evolution of group and network structure. *Social Networks* **1979**, *2*, 235–252. doi:10.1016/0378-8733(79)90016-9.