Entropy Production in Stationary Social Networks

Haye Hinrichsen[1], Tobias Hoßfeld[2], Matthias Hirth[2], and Phuoc Tran-Gia[2]

[1] [1]University of Würzburg, Department of Physics and Astronomy, Am Hubland, 97074 Würzburg, Germany. hinrichsen@physik.uni-wuerzburg.de
[2] [2]University of Würzburg, Institute of Computer Science, Chair of Communication Networks, Am Hubland, 97074 Würzburg, Germany.
[hossfeld,matthias.hirth,trangia]@informatik.uni-wuerzburg.de

Abstract. Completing their initial phase of rapid growth social networks are expected to reach a plateau from where on they are in a statistically stationary state. Such stationary conditions may have different dynamical properties. For example, if each message in a network is followed by a reply in opposite direction, the dynamics is locally balanced. Otherwise, if messages are ignored or forwarded to a different user, one may reach a stationary state with a directed flow of information. To distinguish between the two situations, we propose a quantity called *entropy production* that was introduced in statistical physics as a measure for non-vanishing probability currents in nonequilibrium stationary states. The proposed quantity closes a gap for characterizing social networks. As major contribution, we present a general scheme that allows one to measure the entropy production in arbitrary social networks in which individuals are interacting with each other, e.g. by exchanging messages. The scheme is then applied for a specific example of the R mailing list.

1 Introduction

Due to the rapid growth of social media in the last decade, many theoretical studies have been focused on the growth dynamics of social networks [1]. In such a social network, individuals are connected to each other (e.g. friends in facebook, sender and receiver in mail networks) and there is an interaction between the individuals (e.g. exchanging messages). Hence, beside the network topology, the interaction among invididuals characterize the social network. However, recent observations [2] indicate that many social networks are approaching a plateau of constant size, e.g. due to logistic growth models and resulting upper population bounds like in [3]. In such a matured state, the dynamics of the network are approximately stationary in a statistical sense, meaning that the network topology as well as the probability for receiving and sending messages do not change in the long-term limit.

The dynamics of a stationary state of a network is not uniquely given, rather there is a large variety of possible realizations. For example, the three individuals shown in Fig. 1 may send messages (a) randomly in both directions or (b) in

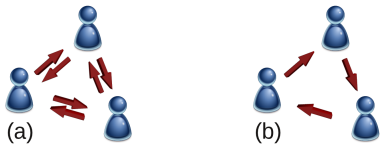


Fig. 1. Example of a stationary network with three users. (a) Each individual sends messages to randomly selected neighbors, leading to a statistically balanced stationary state with vanishing entropy production. (b) The individuals send messages to only one neighbor, generating a stationary but directed flow of information with positive entropy production.

clockwise direction. In both situations the individuals send and receive messages at constant rate, meaning that the network is statistically stationary. However, in the first case the dynamics is locally balanced between pairs of users, while in the second case there is a directed current of messages flowing clockwise through the system.

In the present work we introduce a new type of quantity, called *entropy production* in statistical physics, to characterize the stationary properties of arbitrary social networks. To this end, we associate with each pair of individuals i, j a quantity H_{ij} called entropy, which depends on the number of messages sent from i to j and vice versa. The entropy H_{ij} measures the directionality of the information exchange and vanishes for perfectly balanced communication. Defining the entropy production of a node as the sum over the entropy of all its links, one can identify nodes contributing preferentially to balanced or unidirectional information transfer.

The concept of entropy production requires to make certain assumptions about the dynamics of the network. In particular, we ignore possible correlations between the messages by assuming that the individuals communicate randomly at constant rates. With this assumption each message sent from node i to node j increases the entropy by [4–7]

$$\Delta H_{ij} = \ln w_{ij} - \ln w_{ji} \,, \tag{1}$$

where w_{ij} and w_{ji} are the rates for messages from i to j and in opposite direction, respectively. In physics, this quantity can be interpreted as the minimal entropy produced by a machine that keeps the network running. In computer science this interpretation is irrelevant since realistic networks produce much more entropy in the environment. However, as we will demonstrate in the present work, the entropy production is a useful measure to characterize the stationary properties of the network as, for example, to distinguish the situations (a) and (b) in Fig. 1.

The formula (1) is trivial to evaluate if the rates w_{ij} and w_{ji} are known. However, in realistic networks with data taking over a finite time span T, only the number of messages n_{ij} and n_{ji} exchanged between pairs of nodes are known. Although it is tempting to replace the rates w_{ij} by the relative frequencies n_{ij}/T and to approximate the entropy production by $\Delta H_{ij} = \ln n_{ij} - \ln n_{ji}$, it is easy to see that this approximation would diverge as soon as one of the count numbers

vanishes. Therefore, the paper deals to a large extent with the question how we can reasonably reconstruct the rates from the given number of messages.

The remainder of this paper is structured as follows. Section 2 introduces variables describing the observed data from a measurement campaign of a social network. Further, assumptions on the network dynamics are summarized. Based on that, Section 3 defines the entropy production which is based on the (unknown) message rate between any two individuals of the social network. An estimator of the rates based on the observed measurement data is introduced by means of Bayesian inference. Section 4 presents appropriate choice of the prior distribution for small-world networks, in which the number of messages follows a power-law distribution. Relevant parameters of the prior distribution are calculated which finally allows computing the entropy production. The general scheme is summarized for social networks manifesting small-world characteristics on the number of messages. In Section 5, the general scheme is applied exemplarily to the R mailing list. Section 6 revisits related work in order to show that entropy production fills a gap in characterizing social networks. Finally, Section 7 concludes the work and gives an outlook on next steps in this research direction.

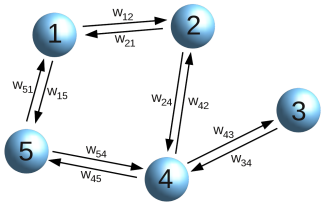
2 Stationary network dynamics

2.1 Observed data

Let us consider a social network of individuals communicating by directed messages (e.g. emails, Twitter or Facebook messages). Suppose that we monitor M messages over a finite time span, recording sender and receiver id's in a file. Such a data set can be represented as a graph of nodes (individuals) connected by directed links (messages) as depicted in Fig. 2.

Enumerating the individuals in the list by $i = 1 \dots N$, let n_{ij} be the number of messages sent from i to j. These numbers constitute a $N \times N$ connectivity matrix which is the starting point for the subsequent analysis.

If at least one message is sent from i to j, the two nodes are said to be connected by a directed link. Obviously, the number of recorded messages M



$$\{w_{ij}\} = \begin{pmatrix} 0 & 0.3 & 0 & 0 & 2.0 \\ 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 3.0 & 0 \\ 0 & 0.7 & 0.2 & 0 & 1.5 \\ 0.2 & 0 & 0 & 1.5 & 0 \end{pmatrix}, \{n_{ij}\} = \begin{pmatrix} 0 & 0 & 0 & 0 & 5 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 4 & 0 \\ 0 & 1 & 1 & 0 & 3 \\ 0 & 0 & 0 & 2 & 0 \end{pmatrix}$$

Fig. 2. Example of a directed social network with $N = 5$ participants. It is assumed that node i sends messages to node j randomly with the rate w_{ij}. Observing the network for a finite time span the number of recorded messages from i to j is n_{ij}. In order to compute the entropy production, one has to estimate the unknown rates w_{ij} from the numbers n_{ij}.

and the total number of directed links L are given by

$$M = \sum_{i,j=1}^{N} n_{ij} \,, \qquad L = \sum_{i,j=1}^{N} (1 - \delta_{0,n_{ij}}) \tag{2}$$

using the Kronecker delta $\delta_{a,b}$, cf. Section 8. Note that $M \geq L$ since two individuals can communicate several times during the observation period.

The statistics of multiple communication is described by the probability distribution

$$P(n) := \frac{\sum_{i,j=1}^{N} \delta_{n,n_{ij}}}{N(N-1)} \tag{3}$$

of the matrix elements n_{ij}. In the present work are particularly interested in social media networks with a small-world topology, where this distribution follows a power law, i.e.

$$P(n) \sim n^{-1-\alpha} \,. \tag{4}$$

2.2 Assumptions on network dynamics

In a realistic social network the messages are causally connected and mutually correlated. As this information is usually not available and requires semantic analysis of the messages, let us consider the messages as uncorrelated instantaneous events which occur randomly like the clicks of a Geiger counter. More specifically, we start with the following assumptions:

- *Stationarity:* We assume that the size of the social network is approximately constant during data taking. This means that the total number N_{tot} of participants in the system is constant. This assumption is e.g. valid for networks following a logistic growth model [3]. Note that N_{tot} may be larger than the actual number of participants N communicating during data taking.
- *Effective rates:* Messages are sent from node i to node j at a constant *rate* (probability per unit time), denoted as $w_{ij} \geq 0$.
- *Reversibility:* If node i is can communicate with note j, node j can also communicate with node i. This assumption is typically true in social networks. Hence if w_{ij} is nonzero, then the rate in opposite direction w_{ji} is also nonzero.

With these assumptions, the average number of communications from i to j is given by

$$\langle n_{ij} \rangle = w_{ij} T \,, \tag{5}$$

where T denotes the observation time.

3 Entropy production

3.1 Definition

As outlined above, each message sent from node i to j produces an entropy of

$$\Delta H_{ij} \;:=\; \ln \frac{w_{ij}}{w_{ji}} \,. \tag{6}$$

Since node i sends n_{ij} messages to node j during the observation period, the total entropy produced by messages $i \to j$ is given by $n_{ij}\Delta H_{ij}$, while messages in opposite direction produce the entropy $n_{ji}\Delta H_{ji}$. Adding the two contributions we obtain the entropy per link

$$H_{ij} \;=\; n_{ij}\Delta H_{ij} + n_{ji}\Delta H_{ji} \;=\; (n_{ij} - n_{ji}) \ln \frac{w_{ij}}{w_{ji}} \,. \tag{7}$$

This entropy is symmetric ($H_{ij} = H_{ji}$) and can equally be attributed to the corresponding nodes, allowing us to define an entropy production per node

$$H_i \;=\; \frac{1}{2} \sum_{j=1}^{N} H_{ij} \tag{8}$$

as well as the entropy production of the total network

$$H \;=\; \sum_i H_i \;=\; \frac{1}{2} \sum_{i,j=1}^{N} H_{ij} \;. \tag{9}$$

3.2 Naïve estimate

The entropy production depends on the message numbers n_{ij} and the rates w_{ij}. While the numbers n_{ij} can be determined directly from the given data, the rates w_{ij} are usually not known in advance. Of course, in the limit of infinite observation time the relative frequencies of messages converge to the corresponding rates, i.e.

$$w_{ij} \;=\; \lim_{T\to\infty} \frac{n_{ij}}{T} \,. \tag{10}$$

For finite observation time the count numbers n_{ij} are scattered around their mean value $\langle n_{ij} \rangle = T w_{ij}$. Therefore it is tempting to approximate the entropy production by replacing the ratio of the rates with the ratio of the relative frequencies, i.e.

$$H_{ij}^{\text{naive}} \;\approx\; (n_{ij} - n_{ji}) \ln \frac{n_{ij}}{n_{ji}} \,. \tag{11}$$

However, this naïve estimator is useless for two reasons. Firstly, the nonlinear logarithm does not commute with the linear average and is thus expected to generate systematic deviations. Secondly, in realistic data sets there may be one-way communications with $n_{ij} > 0$ and $n_{ji} = 0$, producing diverging contributions in

the naïve estimator (11). However, observing no messages in opposite direction does not mean that the actual rate is zero, it only means that the rate is small. In the following we suggest a possible solution to this problem by using standard methods of Bayesian inference, following similar ideas that were recently addressed in a different context [8].

3.3 Bayesian inference

As the messages are assumed to occur randomly like the clicks of a Geiger counter, we expect the number of messages n for a given rate w to be distributed according to the Poisson distribution

$$P(n|w) \;=\; \frac{(Tw)^n e^{-Tw}}{n!}\,, \tag{12}$$

where T is the observation time. But instead of n for given w, we need an estimate of the rate w for given n. According to Bayes formula [9] the corresponding conditional probability distribution is given by the posterior

$$P(w|n) \;=\; \frac{P(n|w)P(w)}{P(n)}\,, \tag{13}$$

where $P(w)$ is the prior distribution and

$$P(n) = \int_0^\infty dw\, P(n|w)P(w) \tag{14}$$

is the normalizing marginal likelihood. The prior distribution expresses our believe how the rates are statistically distributed and introduces an element of ambiguity as will be discussed below. Having chosen an appropriate prior the expectation value $\langle \ln w\rangle$ for given n reads

$$\langle \ln w\rangle_n \;=\; \int_0^\infty dw\, \ln w P(w|n)\,. \tag{15}$$

This allows us to estimate the entropy production of the directed link $i \to j$ by

$$H_{ij} \;\approx\; (n_{ij} - n_{ji})\Big[\langle \ln w\rangle_{n_{ij}} - \langle \ln w\rangle_{n_{ji}}\Big]\,. \tag{16}$$

As we will see, this estimator does not diverge if $n_{ji} = 0$.

4 Small-world networks

4.1 Choice of the prior distribution

The prior should be as much as possible in accordance with the available data. In the example to be discussed below, where we investigate a small-world network with message numbers distributed according to Eq. (4) with an exponent $\alpha > 1$,

it would be natural to postulate a power-law distribution of the rates $P(w) \sim w^{-1-\alpha}$. Since such a distribution can only be normalized with a suitable lower cutoff, a natural choice for the prior would be the inverse gamma distribution

$$P(w) \;=\; \frac{\beta^\alpha w^{-\alpha-1} e^{-\beta/w}}{\Gamma(\alpha)}\,, \tag{17}$$

where the parameter β plays the role of a lower cutoff for the rate w. With this prior distribution the integration can be carried out, giving the posterior

$$P(w|n) \;=\; \frac{(\beta/T)^{\frac{\alpha-n}{2}} w^{n-\alpha-1} e^{-Tw-\frac{\beta}{w}}}{2K_{n-\alpha}(z)}\,, \tag{18}$$

where $K_\nu(z)$ is the modified Bessel function of the second kind and $z = 2\sqrt{\beta T}$. Inserting this result into Eq. (15) we obtain an estimate of $\ln w$ for given n, namely

$$\langle \ln w \rangle_n \;=\; \frac{1}{2}\ln\frac{\beta}{T} + \frac{K^{(1,0)}_{n-\alpha}(z)}{K_{n-\alpha}(z)}\,, \tag{19}$$

where $K^{(1,0)}_\nu(z) = \frac{\partial}{\partial\nu} K_\nu(z)$. The estimator for the entropy production is then given by

$$H_{ij} \;\approx\; (n_{ij} - n_{ji})\Big[\frac{K^{(1,0)}_{n_{ij}-\alpha}(z)}{K_{n_{ij}-\alpha}(z)} - \frac{K^{(1,0)}_{n_{ji}-\alpha}(z)}{K_{n_{ji}-\alpha}(z)}\Big]\,. \tag{20}$$

4.2 Estimating the cutoff parameter $z = 2\sqrt{\beta T}$

Eq. (19) depends on the exponent α, which can be obtained from the distribution of messages per link, and the lower cutoff β, which can be determined as follows. On the one hand, the probability to have no link between two nodes for a given rate w is $1 - P(0|w)$. Therefore, the total number of links L can be estimated by

$$\begin{aligned} L &\approx L_{\rm tot}\int_0^\infty dw\,\Big(1 - P(0|w)\Big)P(w) \\ &= L_{\rm tot}\Big(1 - \frac{2(T\beta)^{\alpha/2}}{\Gamma(\alpha)} K_\alpha(2\sqrt{T\beta})\Big)\,. \end{aligned} \tag{21}$$

Here $L_{tot} = N_{\rm tot}(N_{\rm tot}-1)$ is the unknown total number of potential links in the stationary network which may exceed the actual number of links L established during the finite observation time T.

On the other hand, it is obvious that the total number of messages M can be estimated by

$$M \approx L_{\rm tot}\sum_{n=0}^\infty n\int_0^\infty dw\, P(n|w)P(w) = L_{\rm tot}\frac{T\beta}{\alpha-1}\,. \tag{22}$$

This relation can be used to eliminate $L_{\rm tot}$, turning Eq. (21) into

$$\frac{L\,z^2}{4M(\alpha-1)} \;\approx\; 1 - \frac{2(z/2)^\alpha}{\Gamma(\alpha)} K_\alpha(z)\,. \tag{23}$$

For given M, L, α this approximation interpreted as an equation allows us to numerically determine $z = 2\sqrt{\beta T}$.

4.3 Summary of the procedure

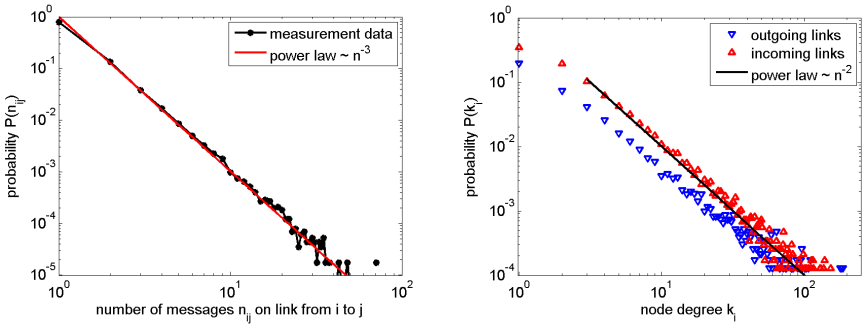
The procedure to calculate the entropy production can be summarized as follows:

1. In the given data set of M messages, identify all participants (nodes) and label them from $1, \ldots, N$.
2. Determine the numbers n_{ij} how often a message is sent from i to j and count the number L of nonzero entries (links) in the matrix $\{n_{ij}\}$.
3. Plot a histogram of the numbers n_{ij}. If it exhibits a power law $P(n) \sim n^{-1-\alpha}$ estimate the exponent α.
4. Solve Eq.(23) numerically for z.
5. Compute the numbers $\chi_n = K^{(1,0)}_{n-\alpha}(z)/K_{n-\alpha}(z)\,.$
6. Associate with each directed link $i \to j$ the entropy production $H_{ij} = (n_{ij} - n_{ji})(\chi_{n_{ij}} - \chi_{n_{ji}})$.
7. Compute H_i and H according to Eqs. (8) and (9).

5 Example: Mailing list archive

To demonstrate the concepts introduced above, we analyzed the mailing lists archive for the programming language **R** [10], recording senders and receivers of all messages over the past 15 years. In this mailing list $N = 23\,462$ individuals (nodes) have exchanged $M = 168\,778$ directed comments (undirected activities like opening a new thread are ignored). The connectivity matrix n_{ij} has $L = 114\,713$ nonzero entries (links). Their statistical distribution shown in Fig. 3 confirms a small-world topology with an exponent $\alpha \approx 2$. Interestingly, the node degree distribution of outgoing and incoming links in Fig. 3(b) seems to exhibit slightly different exponents. A similar phenomenon was observed some time ago in email communication networks [11].

Entropy production per link: Since H_{ij} depends on two integers n_{ij} and n_{ji}, the entropy production of a link produces a discrete set of values. The upper panel of Fig. 4 shows how these values are distributed and how often they occur. As can be seen, the entropy production varies over five orders of magnitude and is distributed irregularly with count numbers ranging from 1 to 10^4.



(a) messages per link (b) links per node

Fig. 3. Degree distributions. (a) Probability $P(n_{ij})$ that a directed link $i \to j$ carries n_{ij} messages. The data is consistent with a power law $P(n_{ij}) \sim n_{ij}^{-3}$, meaning that $\alpha = 2$. (b) Node degree distribution, showing the probability $P(k_i)$ that a node i is connected with k_i outgoing or incoming links. The two data sets display slightly different power laws close to $P(k_i) \sim k_i^{-2}$.

Entropy production per node: Let us now turn to the question how the entropy per node $H_i = \frac{1}{2}\sum_{j=1}^{N} H_{ij}$ is correlated with other properties of the node, in particular with the number of outgoing and incoming messages

$$n_i^{\text{out}} = \sum_j n_{ij}\,, \qquad n_i^{\text{in}} = \sum_j n_{ji}\,. \tag{24}$$

Since the entropy production is expected to grow with the number of messages, it is reasonable to define the node entropy production per message

$$h_i := \frac{H_i}{n_i^{\text{out}} + n_i^{\text{in}}}\,. \tag{25}$$

Fig. 5 shows how the entropy production per node is distributed depending on the number of sent and received messages. As expected, the entropy is minimal if these numbers coincide. Plotting the entropy production of a node versus the

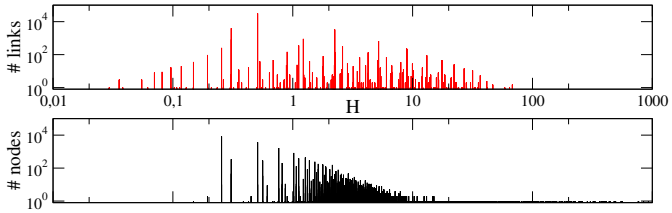


Fig. 4. Upper panel: Histogram of the entropy H_{ij} per link. Lower panel: Histogram of the marginal entropy production H_i per node.

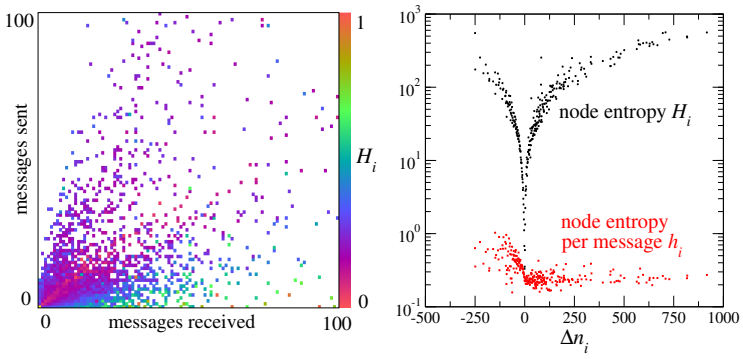


Fig. 5. Left: Average entropy production per node represented by a continuous color scale in a two-dimensional plane spanned by the number of incoming and outgoing messages. Right: Average entropy production of a node (black) and the same data divided by the total number of incoming and outgoing messages (red) as a function of the difference between outgoing and incoming messages.

difference of outgoing and incoming messages

$$\Delta n_i = n_i^{\text{out}} - n_i^{\text{in}} \tag{26}$$

one finds again an asymmetric distribution (see right panel). This indicates that nodes with a large number of outgoing links tend to produce less entropy per message than individuals who preferentially receive messages.

6 Discussions on Related Work

In literature, various measures of the characteristics of complex networks exist. We briefly revisit them in order to show that entropy production fills a gap for measuring the directionality of the information exchange and quantifying the balance of communication or interaction. The quantities introduced in literature analyze mainly the topology of the graph itself by means of the adjacency matrix $\mathcal{A}$ with elements $\mathcal{A}_{ij} = 1 - \delta_{0,n_{ij}}$. Extensions of several quantities exist for weighted networks. In that case, the directed link connecting the nodes i and j are weighted by the message rate w_{ij}. Instead of $\mathcal{A}$, the message rate matrix $\mathcal{W}$ is used. Thereby, beyond the topological effects, the metrics which allow to work on weighted networks give insights into the structure of the message diffusion, too. Those metrics are to be analyzed with the network entropy H for different network topologies and message exchange models $\mathcal{W}$. Future concerns an analysis of those metrics with the entropy production for different network topologies and message exchange models $\mathcal{W}$.

Principal graph characteristics. The basic quantities are the in- and out-degree of nodes corresponding to the number of incoming and outgoing links of nodes.

We observe a strong correlation of the node degree with entropy production for the example of the R mailing list. However, a closer look in the previous section revealed that nodes with a large number of outgoing links tend to produce less entropy per message. Future work investigates for which kind of network topologies and message exchange models those quantities are correlated. Other principal characteristics of nodes are eccentricity and local clustering coefficient. Global network metrics are e.g. radius, diameter, average path length, or assortativity coefficient. Those metrics can be extended to weighted networks and need to be interrelated to entropy production per node and network entropy production, respectively.

Centrality metrics. Centrality metrics quantify the 'importance' of nodes. Different variations exist like degree, (random walk) closeness, information, betweenness, or Eigenvector centrality like PageRank. Considering again the R mailing list, we observed a strong correlation between entropy production per node H_i and e.g. betweenness centrality. Nodes in the social network that have a high probability to occur on a randomly chosen shortest path between two randomly chosen nodes have a high betweenness. Hence, those nodes are also responsible for high entropy production in the network. In contrast, closeness centrality and entropy production revealed no correlation. Closeness measures how fast it will take to spread information from a single node to all other nodes sequentially.

Symmetry measures and entropy measures. Current developmentsintroduce measures of symmetry and their relation with measures like Graph entropy [12]. The concept of Graph entropy is based on a probability distribution on the node set V of the graph [13], $G = \sum_{i=1}^{|V|} p(v_i) \log \frac{1}{p(v_i)}$, but not on the ratio between incoming and outgoing message rates as for entropy production. Hence, graph entropy measures the amount of information within the graph based on $p(v_i)$. Symmetry of complex networks means invariance of adjacency of nodes under the permutations on the node set itself and a symmetry index is defined in [14]. This concept is close to entropy production, however, symmetry measure are defined on network topology only. In a similar way, the symmetry based structure entropy of complex networks [15] quantifies the heteogeneity of a network system based on automorphism partition of the node set into equivalent cells. Thereby, the probability that a node belongs to the cell is used while message rates are not part of this concept. Nevertheless, a comparison of those measures with entropy production is relevant future work.

7 Conclusions

In the current Internet, social networks like Facebook gain more and more popularity and attract millions of users. In a social network the users are connected to each other and as a key feature social media platforms allow interactions between users like exchange of messages. In complex network research, the majority of existing quantities analyze the structural properties of the emerging

network topology and the growth dynamics, respectively. For social networks, however, beyond the network topology the interaction among individuals needs to be characterized. Further, the dynamics of a stationary state of a social network is not uniquely given, rather there is a large variety of possible realizations. Hence, there is a gap in describing dynamical properties of social networks in stationary conditions, i.e., when the network topology as well as the probability for receiving and sending messages do not change in the long-term limit.

Inspired from statistical physics, we introduce a quantity called entropy production to characterize the stationary properties of arbitrary social networks. The entropy production measures the directionality of the information exchange and vanishes for perfectly balanced communication. Defining the entropy production of a node as the sum over the entropy of all its links, one can identify nodes contributing preferentially to balanced or unidirectional information transfer. Hence, entropy production is a valuable measure for link and node analysis and rating and can be used to detect hidden structures and interactions in networks. Since the application of entropy production is not limited to social media network, but can be used for communication networks or interaction graphs in general, it can be applied for a variety of different purposes like anomaly detection [16] but also characterization of traffic flows in the Internet, e.g. for BitTorrent swarms [17]. Future work addresses the application of entropy production to such use cases but also to relate the quantity with centrality or symmetry measures for various network topologies and message exchange models.

References

1. Jin, E.M., Girvan, M., Newman, M.E.J.: Structure of growing social networks. Phys. Rev. E **64** (Sep 2001) 046132
2. Barnes, N.G., Andonian, J.: The 2011 fortune 500 and social media adoption: Have america's largest companies reached a social media plateau? (2011) `http://www.umassd.edu/cmr/socialmedia/2011fortune500/`.
3. Hoßfeld, T., Hirth, M., Tran-Gia, P.: Modeling of Crowdsourcing Platforms and Granularity of Work Organization in Future Internet. In: International Teletraffic Congress (ITC), San Francisco, USA (September 2011)
4. Schnakenberg, J.: Network theory of microscopic and macroscopic behavior of master equation systems. Rev. Mod. Phys. **48** (Oct 1976) 571–585
5. Schreiber, T.: Measuring information transfer. Phys. Rev. Lett. **85** (Jul 2000) 461–464
6. Andrieux, D., Gaspard, P.: Fluctuation theorem and onsager reciprocity relations. J Chem Phys **121**(13) (2004) 6167–74
7. Seifert, U.: Entropy production along a stochastic trajectory and an integral fluctuation theorem. Phys. Rev. Lett. **95** (Jul 2005) 040602
8. Zeerati, S., Jafarpour, F.H., Hinrichsen, H.: Entropy production of nonequilibrium steady states with irreversible transitions. under submission (2012)
9. Box, G.E.P., Tiao, G.C.: Bayesian Inference in Statistical Analysis. John Wiley & Sons, New York (1973 (reprinted in paperback 1992 ISBN: 0-471-57428-7 pbk.))
10. R Mailing Lists: `http://tolstoy.newcastle.edu.au/R/`
11. Ebel, H., Mielsch, L.I., Bornholdt, S.: Scale-free topology of e-mail networks. Phys. Rev. E **66** (Sep 2002) 035103
12. Garrido, A.: Symmetry in complex networks. Symmetry **3**(1) (2011) 1–15
13. Garrido, A.: Classifying entropy measures. Symmetry **3**(3) (2011) 487–502
14. Mowshowitz, A., Dehmer, M.: A symmetry index for graphs. Symmetry: Culture and Science **21**(4) (2010) 321–327
15. Xiao, Y.H., Wu, W.T., Wang, H., Xiong, M., Wang, W.: Symmetry-based structure entropy of complex networks. Physica A: Statistical Mechanics and its Applications **387**(11) (2008) 2611–2619
16. Bilgin, C., Yener, B.: Dynamic network evolution: Models, clustering, anomaly detection. Technical report, Rensselaer University, NY (2010)
17. Hoßfeld, T., Lehrieder, F., Hock, D., Oechsner, S., Despotovic, Z., Kellerer, W., Michel, M.: Characterization of BitTorrent Swarms and their Distribution in the Internet. Computer Networks **55(5)** (April 2011)

8 Appendix: Notion of variables frequently used

For the sake of readability, the appendix is included for review only and will be removed in the camera-ready version.

	variables describing the measurement data
T	measurement period over which messages between individuals are recorded
n_{ij}	number of messages sent from i to j during time T
N	total number of individuals, i.e. nodes in the graph, communicating during T
M	total number of recorded messages, i.e. directed link, $M = \sum_{i,j=1}^{N} n_{ij}$
$\delta_{a,b}$	Kronecker delta defined by $\delta_{a,b} = \begin{cases} 1 & a = b \\ 0 & a \neq b \end{cases}$
L	total number of directed links, $L = \sum_{i,j=1}^{N} 1 - \delta_{0,n_{ij}}$
n_i^{out}	number of outgoing messages from node i, $n_i^{\text{out}} = \sum_{j=1}^{N} n_{ij}$
n_i^{in}	number of incoming messages to node i, $n_i^{\text{in}} = \sum_{j=1}^{N} n_{ji}$
Δn_i	difference of outgoing and incoming messages of node i
$P(n)$	probability that n messages are sent on a link, $P(n) = \sum_{i,j=1}^{N} \delta_{n,n_{ij}} / N(N-1)$
$\mathcal{A}$	adjacency matrix with matrix elements $\mathcal{A}_{ij} = 1 - \delta_{0,n_{ij}}$
	variables describing entropy production
w_{ij}	message rate from i to j estimated by measured n_{ij} over T
$\mathcal{W}$	rate matrix with matrix elements $\mathcal{W}_{ij} = w_{ij}$
ΔH_{ij}	amount of entropy increased for each message sent from i to j, $\Delta H_{ij} = \ln \frac{w_{ij}}{w_{ji}}$
H_{ij}	entropy per link, $H_{ij} = (n_{ij} - n_{ji}) \ln \frac{w_{ij}}{w_{ji}}$
H_i	entropy production per node i, $H_i = \frac{1}{2} \sum_{j=1}^{N} H_{ij}$
H	entropy production of total network, $H = \sum_{i=1}^{N} H_i$
h_i	node entropy production per message, $h_i = \frac{H_i}{n_i^{\text{out}} + n_i^{\text{in}}}$
	variables for estimating message rates
$P(w\|n)$	posterior distribution of message rates w conditional on observed messages n
$P(w)$	prior distribution of message rates; assumed to follow a power law in social networks with $P(w) \sim w^{-1-\alpha}$; normalization with suitable lower cutoff leads to inverse gamma distribution $P(w) = \frac{\beta^\alpha}{\Gamma(\alpha)} w^{-\alpha-1} e^{-\beta/w}$
α	shape parameter of the inverse gamma distribution
β	lower cutoff parameter for the rate w concerning inverse gamma distribution
$P(n)$	normalizing marginal likelihood
$\langle \ln w \rangle_n$	expectation value for given n, $\langle \ln w \rangle_n = \int_0^\infty dw \, \ln w P(w\|n)$
$K_\nu(z)$	modified Bessel function of the second kind and $z = 2\sqrt{\beta T}$