Separating temporal and topological effects in walk-based network centrality

Ewan R. Colman and Nathaniel Charlton

Separating temporal and topological effects in walk-based network centrality

Ewan R. Colman

Department of Biology, Georgetown University, Washington, DC 20057, U.S.A∗

Nathaniel Charlton

CountingLab Ltd., Reading, United Kingdom, RG6 6AX

The recently introduced concept of dynamic communicability is a valuable tool for ranking the importance of nodes in a temporal network. Two metrics, broadcast score and receive score, were introduced to measure the centrality of a node with respect to a model of contagion based on time-respecting walks. This article examines the temporal and structural factors influencing these metrics by considering a versatile stochastic temporal network model. We analytically derive formulae to accurately predict the expectation of the broadcast and receive scores when one or more columns in a temporal edge-list are shuffled. These methods are then applied to two publicly available data-sets and we quantify how much the centrality of each individual depends on structural or temporal influences. From our analysis we highlight two practical contributions: a way to control for temporal variation when computing dynamic communicability, and the conclusion that the broadcast and receive scores can, under a range of circumstances, be replaced by the row and column sums of the matrix exponential of a weighted adjacency matrix given by the data.

PACS numbers: 64.60.aq 89.65.-s

I. INTRODUCTION

Epidemics, viral marketing, cultural diffusion, the distribution of food in ant colonies, and the flow of information within the human brain, are amongst a growing number of applications of network theory which currently reside at the forefront of modern science [1–5]. Advances in technology continue to promote the accumulation of data, providing an optimistic light in the quest to understand these hugely complex systems. The task then, for researchers across a range of disciplines, is to find optimal ways to measure, model, analyze, and present the vast information at their disposal.

Network theory has proved to be an invaluable resource to exploit data on a large scale. Its great utility comes partly from the its ability to translate problems into a language independent of the particular subject of study. Hence, a "node" can represent entities as diverse as a human, a protein or a word [6–8]. "Edges" can represent any sort of interaction between the nodes, and concepts such as percolation, diffusion, paths and walks can all serve as models for various processes observed in the real world.

It is remarkable whenever the methods developed for the analysis of one subject matter are applied to seemingly unrelated problems. This occurs frequently when networks are involved. For example, the preferential attachment model can explain the distribution of citations in scientific literature as well as the distribution of popularity in a social network [9, 10], the PageRank algorithm was developed to rank websites but can also measure the risk of cancer in humans [11]. These universalities motivate us to search for ways to measure networks and classify them by their properties; if we have a good description of the network, then we have potentially described a part of the "real world" which we would like to understand, moreover, we also have the entirety of past research and all the accompanying tools developed to help attack the problem.

A. Motivation for "dynamic communicability"

Transmissible disease is possibly the best example to demonstrate the versatility of network analysis. Ultimately the theoretical considerations of network epidemiology involve nodes, edges and some knowledge of the disease itself such as the transmission probability, recovery rate and so on. Transmission could occur from one person to another, from one location to another (e.g. connected by air travel), or between species, but in each case the models employed remain well within the confines of the network framework [12–14]. This also extends to computer viruses [15], Twitter hashtags and internet memes [16, 17], and possibly even cultural transmission on an archeological time-scale [18]. Clearly there is much to be gained from having a grounded understanding of how things spread through a network regardless of what that particular network represents.

The work we present here concerns a scenario where we are given a database containing a set of distinct individuals, a set of pairwise interactions, and the exact time at which each interaction happened (see Fig.1a). Additionally it is assumed that some transmissible agent was, or potentially could have been, spreading through the network. A practical question which often arises is: "which node is potentially the most significant when it comes to the spread of a transmissible agent?".

To find the most influential spreader, given data of

∗ ec975@georgetown.edu

past interactions, there are several options to consider: the simplest method would be to find the individual with the highest node degree (this could be defined as either number of interactions that person had, or the number of people with whom they interacted). Alternatively we could use global network properties such as the betweenness centrality or closeness centrality of a node, both of which are defined on temporal networks [19]. The most extensive approach currently being used is to build a computational model of the process, adding as many factors into the model as one sees fit; where uncertainty is present, random variables can be used; and the centrality of an individual can be computed by running the model repeatedly and counting the proportion of simulations in which they are infected [12, 20].

Dynamic communicability, which was introduced in [21] and is described in detail here in Section II, offers a balance between the approach of modeling an epidemic-like process on a network, and simply measuring the size and shape of a network. Here we determine the influence of a node by counting the number of time-respecting walks that began at the node in question. In essence, we are using a model which assumes that a transmissible agent moves from one node to another at the exact time that an interaction takes place (which is known from the data) and with a given transmission probability. The fact that it is a walk (as opposed to a path) means that the agent can revisit previously infected nodes. Assuming this, and supposing that the pathogen is administered at node i, the broadcast score of i tells us how large the expected outbreak will be. Supposing the pathogen is administered to a random unknown node, the receive score of i tells us how likely that pathogen is to reach i.

B. Separating dependencies

In this paper we interrogate the two dynamic communicability metrics: broadcast score and its opposite, receive score. Through theoretical approaches we will examine how these centrality measures respond to different temporal network structures. Further, we derive methods to deconstruct the dynamic communicability measures into "time dependent" and "structure dependent" components. The formulae we derive achieve the same result as "shuffling" (randomly permuting) either the structural or temporal columns of the temporal edge-list respectively. This is an increasingly common technique used to determine the importance of various relationships within a database [20, 22–24]. Here we employ this technique to unpick, from the information available, the factors most relevant to determining the outcome of a contagion-like process.

The following section explains in detail the dynamic communicability metrics. In Section III we describe a stochastic model which can be tuned to reproduce various properties of the data. The main results from the model are a set of "shortcut formulae" for decompos-

ing the dynamic communicability metrics into time dependent and structure dependent elements in an efficient way. We demonstrate these results on two publicly available data sets, which are described in Section IV, and the results are presented in Section V. Section VI summarizes the findings from this work which we consider most significant.

II. DEFINITIONS OF "BROADCAST SCORE" AND "RECEIVE SCORE"

Communicability, as introduced in [25], is a measure of centrality based on the concept of "walking" on a network. A walk is any sequence of nodes in which one entry may only follow another if there is an edge in the network which connects them (if the network is directed then consecutive entries must follow the direction of the edge). The extension to temporal networks, in which edges exist only at specified temporal instances, was introduced in [21] and further developed in [26] and [27]. When dealing with temporal edges, we consider node sequences in which consecutive nodes are connected by an edge and, additionally, the time of that edge is later than (or at the same time as) its predecessor. These are referred to as "time-respecting" walks.

Based on this premise it is possible to quantify the relationship between any two nodes: the "dynamic communicability" from node i to node j, denoted $Q_{i,j}$, is a measure of the relative likelihood that a random walker injected into the network at i will eventually pass through j. If we let $\theta_{i,j}^{(k)}$ be the number of time-respecting walks of length k that begin at i and end at j, then

$$Q_{i,j} = \sum_{k=0}^{\infty} \alpha^k \theta_{i,j}^{(k)}. \tag{1}$$

The value α here is analogous to the probability of transmission (across an edge) in an epidemic spreading process. When chosen to be sufficiently small, it ensures that long walks are discounted heavily while short walks contribute more to the dynamic communicability metric. Several alternative approaches have been proposed as ways to down-weight walks based on their length. The original communicability defined on a static network discounts walks of length k by dividing their number by $k!$, whereas when the temporal equivalent was introduced the exponential discounting shown in Eq.(1) was used. The measure introduced in [26] combines both. The measure introduced in [27] extends Eq.(1) by additionally discounting walks according to their duration in time.

For the first centrality measure, known as the "Broadcast score" of a node i, we compute the sum of all the discounted walks that begin at i ($b_i = \sum_{j \in N} Q_{i,j}$). Similarly, to compute the second centrality measure, known as the "Receive score" of a node i, we sum all of the discounted walks that end at i ($c_i = \sum_{j \in N} Q_{j,i}$).

(a) The left panel shows the network at each time-step (above) and its corresponding adjacency matrix (below). On the right the same information is represented as a list of temporal edges. Shown also are the different possible ways to randomize (or shuffle) the columns. Notice that simultaneously shuffling any two columns yields the completely shuffled edge-list shown in (iv).

(b) Each marker corresponds to a node in the example network. Each node is given a rank according to its broadcast score (left) and receive score (right). These rankings are plotted against the outgoing and incoming degree ranks respectively. The diagonal line divides the nodes into those that acheive higher broadcast (or receive) scores than expected, and those that are lower.

(c) The expectation values for each shuffling are calculated and the corresponding rank is plotted (we have chosen only to consider the target shuffling for broadcast score and source shuffling for receive score). The actual scores are shown by the darkness of the markers.
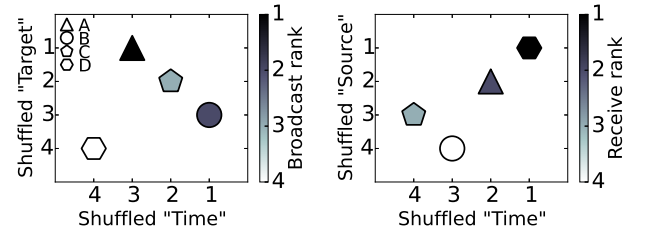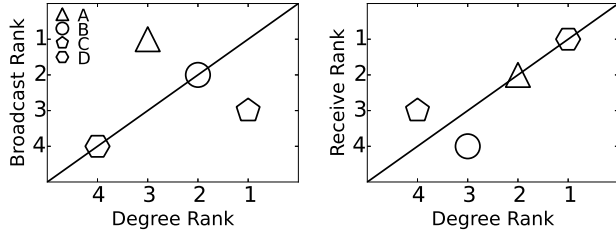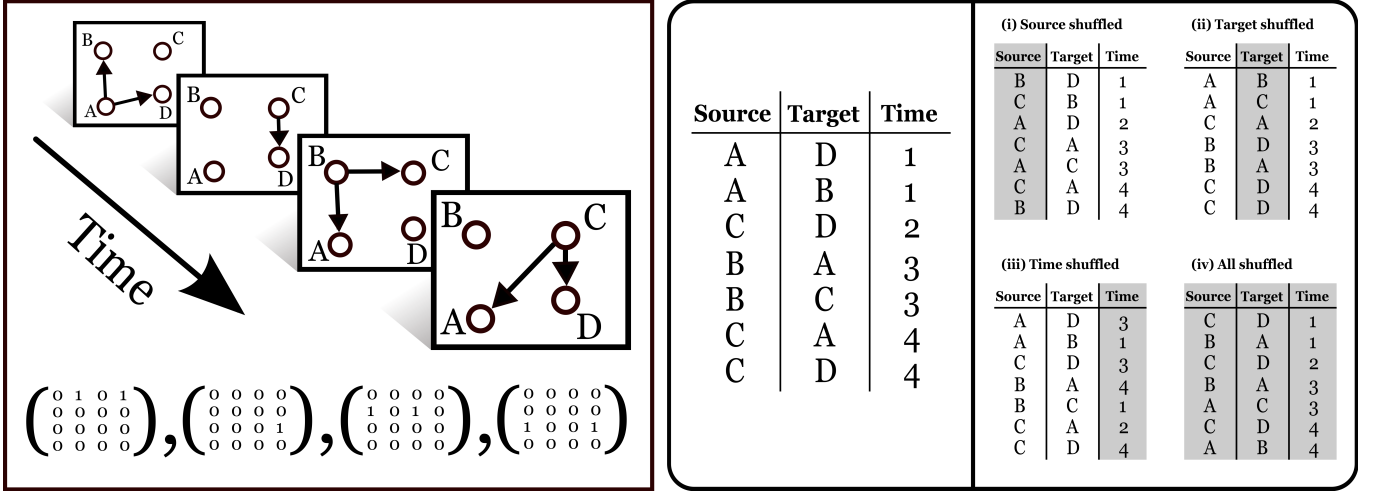
FIG. 1. A simple example of a directed temporal network. This example has been designed to illustrate the core concepts of this work. In Fig.(1b) it is apparent that the dynamic communicability of a node is not necessarily determined by its overall activity. Fig.(1c) demonstrates how the dynamic communicability metrics can be broken down into temporal and structural elements. We apply the same visualization method to two real-world data-sets in Figs. (4) and (5).

III. THE MODEL

We use a simple yet versatile stochastic model to generate temporal networks. The parameters of the model can be manipulated to create synthetic data with properties similar to a wide range of temporal networks including those observed in many real interactive systems. Let there be N nodes. The model proceeds over a series of discreet time-steps $\tau \in \{t_0, t_0 + 1, ..., t_{\text{end}}\}$ by the following rule:

At time τ, with probability $\rho_{i,j}(\tau)$, a directed edge exists from node i to node j.

The adjacency matrix at time τ, A_τ, will have a 1 in location i, j with probability $\rho_{i,j}(\tau)$ and be 0 otherwise (it might often be the case that A_τ will be a matrix of zeros). The dynamic communicability matrix, as introduced in [21], over the sample (starting at t_0 and ending at t_{end})

is given in general by

$$Q = (I - \alpha A_{t_0})^{-1}(I - \alpha A_{t_0+1})^{-1} \ldots (I - \alpha A_{t_{\text{end}}})^{-1} \quad (2)$$

where I denotes the identity matrix. But, as suggested in [28], we do not want to count paths that take multiple moves in a single time-step, so we will instead look at the variant definition

$$Q = (I + \alpha A_{t_0})(I + \alpha A_{t_0+1})...(I + \alpha A_{t_{\text{end}}}). \quad (3)$$

Eqns. (2) and (3) are equivalent when $A_\tau^2 = 0$ for all τ (as this is the only way $(I - \alpha A_\tau)^{-1} = (I + \alpha A_\tau)$ can be true) i.e. provided that no walks of length 2 ever exist within a single time slice. Under these conditions our analysis also applies to the version of dynamic communicability defined in [26], where $Q = \exp(\alpha A_{t_0})\exp(\alpha A_{t_0+1})...\exp(\alpha A_{t_{\text{end}}})$, since $\exp(\alpha A) = (I + \alpha A)$ when $A^2 = 0$.

The time-dependent matrix

$$P(\tau) = \begin{pmatrix} \rho_{1,1}(\tau) & \rho_{1,2}(\tau) & \cdots & \rho_{1,N}(\tau) \\ \rho_{2,1}(\tau) & \rho_{2,2}(\tau) & \cdots & \rho_{2,N}(\tau) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{N,1}(\tau) & \rho_{N,2}(\tau) & \cdots & \rho_{N,N}(\tau) \end{pmatrix}, \quad (4)$$

to a large extent, describes the entire structure of the network and its evolution over time. Our approach to exploring dynamic communicability of networks generated by this model involves considering the various forms that P can take; then examining the expectation of Q as we iteratively increase the number of terms on the right hand side of Eq.(3). The following analysis requires that the values contained in P are small enough that the probability of generating a matrix containing a path of length 2 or more is negligible. The model is therefore more applicable to temporally highly resolved data-sets as opposed to those in which a relatively small number of temporal instances are recorded.

A. Receive score

If we think about constructing Q iteratively i.e. starting at time t_0 with $Q_0 = (I + \alpha A_{t_0})$, then multiplying on the right by $(I + \alpha A_{t_0+1})$, then again by the next term, then the next etc., then

$$Q_t = Q_{t-1} \times (I + \alpha A_{t_0+t}), \quad (5)$$

where t indexes the number of times the iteration has been performed. After $t = t_{\text{end}} - t_0$ iterations we have the desired $Q_t = Q$. The effect of one iteration can be seen on a 3×3 example:

$$\begin{pmatrix} q_{1,1} & q_{1,2} & q_{1,3} \\ q_{2,1} & q_{2,2} & q_{2,3} \\ q_{3,1} & q_{3,2} & q_{3,3} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \alpha \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} q_{1,1} & q_{1,2} & q_{1,3}+\alpha q_{1,2} \\ q_{2,1} & q_{2,2} & q_{2,3}+\alpha q_{2,2} \\ q_{3,1} & q_{3,2} & q_{3,3}+\alpha q_{3,2} \end{pmatrix}. \quad (6)$$

In general, provided $A_\tau^2 = 0$ for all $t_0 < \tau \leq t_{\text{end}}$, if the i,jth entry of A_τ is 1 then the ith column is multiplied by α and added to the jth column. Since the receive score after t iterations is equal to the (row) vector of column sums of Q_t,

$$\mathbf{c}(t) = (c_1(t), c_2(t), ..., c_N(t)), \quad (7)$$

we can describe its evolution as t increases as follows: at each iteration choose i and j with probability $\rho_{i,j}(t_0+t)$ and update by setting

$$\mathbf{c}(t+1) = (c_1(t), c_2(t), ..., c_j(t)+\alpha c_i(t), ..., c_N(t)). \quad (8)$$

In matrix notation this is

$$\mathbf{c}(t+1) = \mathbf{c}(t)(I + \alpha A_{t_0+t}). \quad (9)$$

B. Expectation of receive score

The receive score is dependent on t. To examine this dependence, we focus on the expectation of $c_i(t)$, denoted $\hat{c}_i(t)$, which is computed by taking the mean over many networks generated by the described model for some given P. For analytical considerations we assume that all of the $c_i(t)$ are well approximated by their mean. A similar approach is found in [29]. The growth of $\hat{c}_i(t)$ is then described by

$$\hat{c}_i(t+1) = \hat{c}_i(t) + \alpha \sum_{j=1}^N \rho_{j,i}(t_0+t)\hat{c}_j(t). \quad (10)$$

The right hand side here equation sums over all possible changes that can happen to c_j and their associated probabilities. This is equivalent to replacing A_{t_0+t} in Eq.(9) with the expectation of A_{t_0+t}, which happens to be $P(t_0+t)$. We have

$$\hat{\mathbf{c}}(t+1) = \hat{\mathbf{c}}(t)[I + \alpha P(t_0+t)]. \quad (11)$$

For large time-scales, we can say that $\hat{c}_j(t+1) - \hat{c}_j(t) \approx \partial\hat{c}_j/\partial t$, giving

$$\frac{\partial\hat{\mathbf{c}}(t)}{\partial t} = \alpha\hat{\mathbf{c}}(t)P(t_0+t). \quad (12)$$

An almost identical derivation can be performed to find a similar expression for $\hat{\mathbf{b}}$. In this case, instead of starting the iterative process at t_0 and multiplying on the right, as in Eq.(5), we start at time t_{end} with $Q_0 = (I + \alpha A_{t_{\text{end}}})$ and iterate by multiplying on the left, i.e $Q_t = (I + \alpha A_{t_{\text{end}}-t}) \times Q_{t-1}$. Following similar steps we arrive at

$$\frac{\partial\hat{\mathbf{b}}(t)}{\partial t} = \alpha P(t_{\text{end}}-t)\hat{\mathbf{b}}(t) \quad (13)$$

where $\hat{\mathbf{b}}(t)$ is a column vector of the expectation of the broadcast scores. Our theoretical results stem from these two equations, solutions can be found for various forms of $P(\tau)$, here we mention a few simple cases.

C. Time-independent P matrix

When P is a constant matrix, the (well known) general solution to Eq.(12) is

$$\hat{\mathbf{c}}(t) = \hat{\mathbf{c}}(t_0)e^{\alpha P t} \quad (14)$$

where

$$e^{\alpha P t} = \sum_{k=0}^\infty \frac{1}{k!}(\alpha P t)^k. \quad (15)$$

1. Equivalence to shuffling the time column

Consider a temporal edge-list where the "time" column has been shuffled as shown in part (iii) of Fig.(1a). The overall number of interaction events between each pair of nodes is unchanged, however each of these events now occurs at some random point in time. The time-series of interaction events from node i to node j can be modeled by a Bernoulli process, i.e. at each discreet time-step there is a fixed probability that an edge from i to j will exist. If we have a sufficiently large amount of data then the matrix of these time-independent probabilities, which happens to be P, can be approximated easily as we show in this section. The above result can then be used to predict the dynamic communicability metrics of the time-shuffled edge-list.

We can infer P from the data by constructing a weighted adjacency matrix W where $W_{i,j}$ is the total number of times each edge appears in the temporal edge-list. To infer a time-independent probability $\rho_{i,j}$ that an edge exists at time τ (for any $t_0 \leq \tau < t_{\text{end}}$) we normalize by the number of time steps in the sample:

$$\rho_{i,j} = \frac{W_{i,j}}{t_{\text{end}} - t_0}. \tag{16}$$

Since $t_0 - 1$ lies outside the time for which data is sampled, A_{t_0-1} is a zero matrix, giving $\hat{\mathbf{c}}(t_0 - 1) = \mathbf{1}$ where $\mathbf{1}$ is a row vector of length N and all entries are 1. Substituting the P matrix associated with Eq.(16) into Eq.(14), and into the equivalent result for $\hat{\mathbf{b}}$, we arrive at the concise formulae for computing the expectation of the broadcast and receive scores of a time-shuffled network,

$$\hat{\mathbf{b}} = e^{\alpha W} \mathbf{1} \tag{17}$$

and

$$\hat{\mathbf{c}} = \mathbf{1}^T e^{\alpha W} \tag{18}$$

respectively. A very fast open-source algorithm for solving the matrix exponential for large matrices has recently been developed [30]. Applying this method gives a prediction for the outcome of averaging a large number of shuffled temporal edge-lists where the "time" column has been shuffled. The comparison between the prediction and the actual shuffled data is shown in Fig.(2).

2. Heterogenous "send" and "receive" model

Consider a temporal edge list for which all three columns have been shuffled as in part (iv) of Fig.(1a). While much of the relational information will be lost, the number of times each node is found in the "source" column will be unchanged and therefore the outgoing degree of each node is retained, similarly the incoming degree is unchanged by the shuffling of the "Target" column. This process bears much resemblance to the configuration model of [31] in which each node has a given degree

but the pairwise connections are randomized. Related models, which replace the exact degree sequence with a sequence of fitness variables (giving the propensity of each node to attract edges), have been studied [32]; this happens to be a case where Eqs.(17) and (18) can be solved analytically.

Let s_i be the probability that node i has an outgoing edge in any given time-step (we have chosen the letter s as this represents the 'sending' of information), and let r_i be the probability that i has an incoming edge in any given time-step (r to represent the 'receiving' of information). With the vector notation, $\mathbf{s} = [s_1, s_2, \ldots, s_N]^T$ and $\mathbf{r} = [r_1, r_2, \ldots, r_N]^T$, we have

$$P = \mathbf{s}\mathbf{r}^T. \tag{19}$$

We add the condition that $\sum_i s_i = \sum_i r_i = 1$ then the expected number of edges per time-step is 1 (meaning that when comparing to data we can treat t as the total number of interactions). Under these conditions the solution to Eq.(12) (see Appendix VII A) is

$$\hat{\mathbf{c}}(t) = \mathbf{1} + \frac{e^{\alpha[\mathbf{r}^T\mathbf{s}]t} - 1}{\mathbf{r}^T\mathbf{s}} \mathbf{r}^T. \tag{20}$$

Two main conclusions come from this result: firstly, the receive score of a node is proportional to its propensity to attract incoming edges (for broadcast score it is the outgoing edges, see Appendix VII A). Second, as the sample size t increases the score increases exponentially.

D. Time-dependent P matrix

A general solution to Eq.(12) for any $P(\tau)$ does not exist, we instead incorporate a limited amount of temporal information by expanding the "send" and "receive" model of the previous section. Suppose we have the model from Section III C 2 with the modification that the "receive" vector $\mathbf{r}$ is now a function of time, say $\mathbf{r}(\tau)$, then Eq.(12) reduces to

$$\frac{\partial \hat{\mathbf{c}}(t)}{\partial t} = \alpha \mathbf{r}(t_0 + t) e^{\alpha \mathbf{s}^T \int_0^t \mathbf{r}(t_0+t')dt'} \tag{21}$$

(see Appendix VII B). Eq.(21) allows us to examine special cases where the order in which messages are sent affects the receive score of each node.

1. Simple time-dependent example

Before we derive a result applicable to real-world data, we introduce a simple example to provide some intuition for the time-dependence. We consider the case where each node is active only once during the duration of the sample. Suppose node i receives r_i edges at time τ_i. We can write the corresponding r vector using the Dirac δ:

$$r_i(\tau) = r_i\delta(\tau - \tau_i). \tag{22}$$

The justification for this choice of $\mathbf{r}(\tau)$ is that the expected number of messages received by i over some interval will be r_i if the time interval includes τ_i. For convenience we suppose, without loss of generality, that $\tau_i = i$ for all $i \in 1, 2, ..., N$. In Appendix VII C Eq.(20) is solved with this form of $\mathbf{r}(\tau)$ to get

$$\hat{c}_i(N) = 1 + \alpha r_i \exp\left(\alpha \sum_{j=0}^{i} r_j s_j\right). \qquad (23)$$

This result shows that nodes which interact later in the sample will have, on average, exponentially higher receive scores. In a similar way, it can be shown that a node which acts earlier in the sample has an exponentially higher broadcast score.

2. Incorporating empirical data

Suppose that for each node i we know the time of every received edge but do not know where the edge originated from (this corresponds to the source shuffled network). We can achieve this by choosing

$$r_i(\tau) = \sum_{k \in K_i} \delta(\tau - \tau_i^{(k)}) \qquad (24)$$

where K_i is the set of edges for which i is the target and $\tau_i^{(k)}$ is the time at which edge k was present. More important, however, is the function $R_i(\tau)$ which we define as the number of messages that have been received by i between t_0 and τ, and can be expressed as

$$R_i(\tau) = \int_{t_0}^{\tau} r_i(t')dt'. \qquad (25)$$

To achieve the correct normalization (for the expectation of the total number of edges to agree with the data) we choose s_i to be the probability that any given edge is sent from i, this is inferred using

$$s_i = \frac{\sum_j W_{i,j}}{\sum_{i,j} W_{i,j}}. \qquad (26)$$

The solution to Eq.(12), which we derive in Appendix VII D, is

$$\hat{c}_i = 1 + \alpha \sum_{k \in K_i} \exp\left(\alpha \sum_{j \in N} s_j R_j(\tau_i^{(k)})\right). \qquad (27)$$

This formula predicts the average of the receive score over many networks generated by shuffling the Source column in the original data. The analytical prediction and average shuffling results are shown in Fig.(2). In our data analysis we also use an equivalent formula to predict the outcome of shuffling the target column and
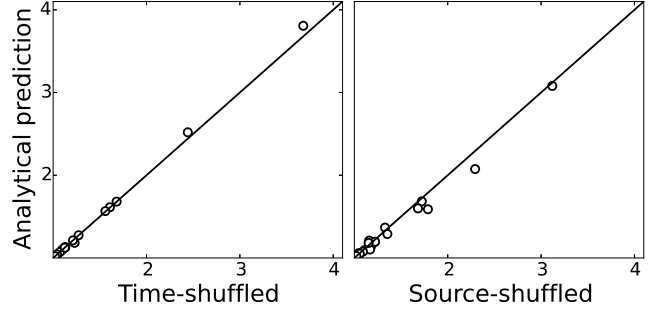

FIG. 2. A demonstration of the accuracy of the derived formulae using a sample of 23 nodes from the Enron data-set (all of which have at least one outgoing edge within the sample) and a total of 312 emails. Each marker represents an employee. In both plots the x-axis shows the receive score computed by shuffling one column of the edge-list, as shown in Fig.(1a), and averaged over 100 shuffles. In the left hand plot the time column was subjected to shuffling and y-axis shows the receive score as predicted by Eq.(18), in the right the target column was subjected to shuffling and y-axis shows the receive score as predicted by Eq.(27). $\alpha = 0.02$.

calculating the broadcast score. The derivation is similar to that of Eq.(27). We get

$$\hat{b}_i = 1 + \alpha \sum_{k \in K_i} \exp\left(\alpha \sum_{j \in N} r_j S_j(\tau_i^{(k)})\right) \qquad (28)$$

where $S_j(\tau)$ is the number of messages that have been received by i between τ and t_{end}, r_j is the time-independent probability that j receives a message in any given time-step.

IV. DATA

A. Enron

We downloaded the entire Enron email corpus that was made publicly available during an investigation by the Federal Energy Regulatory Commission into the events leading to its bankruptcy [33]. The data contains the mailing history of 150 Enron employees between 1999 and 2003. A folder exists for each of the named employees, each of which contain a number sub-folders, and each subfolder contains a number of text files; the text files contain the emails themselves and some meta-data. The naming of the folders is not consistent across employees; most sent emails belong to a folder labelled "sent", "sent email", or something similar but there are also many exceptions. A consistent format was found across all the text files with the time-stamp located on the first line, the "From" field appearing on the second, and the "To" field starting on the third line and often extending over several lines where emails have been sent to multiple recipients.

We crawled every text file within sub-folders named "sent", "sent_items" and "_sent_mail', reading the specific lines which correspond to the "From" field, the "To" field and the time-stamp. Within the "From" and "To" lines we found all substrings which resemble a distinct email address i.e. bound on either side by blank spaces and contain the "@" symbol. From these data we constructed a temporal edge list of the form shown in Fig.(1a) where the node IDs are email addresses. Multiple edges were created for emails with multiple recipients. In several cases the email addresses found in the "From" field, across the emails of an individual employee, would not always be identical. Usually this was because of the use of email aliases although on a small number of occasions this was clearly not the case. At our own discretion, we replaced the node ID of all aliases relating to an employee with a single node ID.

Many of the emails were sent to addresses outside of the corporation, these were removed from our data. We also found that some employees in the data-set had very little or no activity; we therefore reduced the sample to only those who have both sent, and received, at least one email to other users within the sample. After trimming, the network has 141 nodes and a total of $21,303$ temporal edges.

We also incorporated information regarding the roles of each employee according to enron.org [34]. The following abbreviations have been used for the legend in Fig.(7b): EMP=employee, TRA=trader, LAW=lawyer, MAN=manager, DIR=director, VP=vice president, MD=managing director, PRE=president, CEO=chief executive, ???=unknown.

The sample of emails we have chosen to use is by no means complete, however, it is our belief that the methods used to sample this data avoid introducing any biases which would compromise the results we present.

B. Sociopatterns hospital ward

We downloaded the Hospital ward dynamic contact network from the Sociopatterns website (refer to [35] for details). The data was collected using proximity sensors attached to each participant. In the original data, every instance (instances are recorded every 20 seconds) in which two participants are "interacting" (i.e. within a given proximity of each other) is presented in a temporal edge list of the form shown in Fig.(1a). Consequently, interactions which occur for a prolonged duration appear in the data multiple times so we performed the following reduction: where the same pair of participants were found to be interacting on multiple consecutive time-steps, all but one of the corresponding rows in the edge list were removed, leaving only the first of such instances. For each remaining row we create two edges in the processed temporal edge-list, one in each direction between the pair of participants interacting, both edges have the same time-stamp. Our analysis therefore considers transmission to

occur at the first moment an interaction begins and does not depend on its duration. After processing, the network has 75 nodes and a total of $28,076$ temporal edges.

C. Algorithms

Much of the related literature formulates the problem of computing a dynamic communicability matrix using a series of linear algebra operations [21]. This approach utilizes the adjacency matrix for the network at each time step (see Fig.(1a)) and assumes that within each time-slice the hypothetical random walker can traverse edges instantaneously, i.e. without requiring that time move forward for them to perform the movement. Consequently, if there is any cycle within a single time-slice (including for example an edge from i to j and another from j to i) then there will be paths of infinite length, meaning that α must be restricted to a particular range of values to guarantee convergence [36].

In this work we remove the assumption that a walk can traverse more than one edge per time slice (as suggested in [28]). Moreover, we suggest the following recursive approach to computing the dynamic communicability metrics which avoids the need to perform any matrix operations.

Suppose we have a network G with each temporal edge denoted by a triple (i, j, t) where i is the source node, j is the target node and t is the time. Rewriting Eq.(9) with this notation we have

$$c_i(\tau + 1) = c_i(\tau) + \alpha \sum_{(i,j,t) \in G} c_j(t) \qquad (29)$$

with $c_i(t_0 - 1) = 1$. Then the receive score for node i computed between time t_0 and t_{end} is given by

$$c_i = c_i(t_{\text{end}}). \qquad (30)$$

Similarly, for the broadcast score we have

$$b_i(\tau - 1) = b_i(\tau) + \alpha \sum_{(i,j,t) \in G} b_j(t) \qquad (31)$$

with $b_i(t_{end} + 1) = 1$. Then the broadcast score for node i computed between time t_0 and t_{end} is given by

$$b_i = b_i(t_0). \qquad (32)$$

If we were to first compute the vector $c_i(t_0)$ for all the nodes i, then commit these values to memory, then compute $c_i(t_{\text{end}} + 1)$ for all i, and continue in this fashion, then the addition operations we perform are precisely the same as those performed in the established matrix multiplication method [28]. The advantage of this implementation, however, is that the score for a single node can be computed lazily, that is, without wasting unnecessary time. (It is important, when using this method, to use memoization to avoid repeating a large number of calls to the functions c_i and b_i). Computationally we can be certain that these algorithms are at least as fast as the current alternatives.
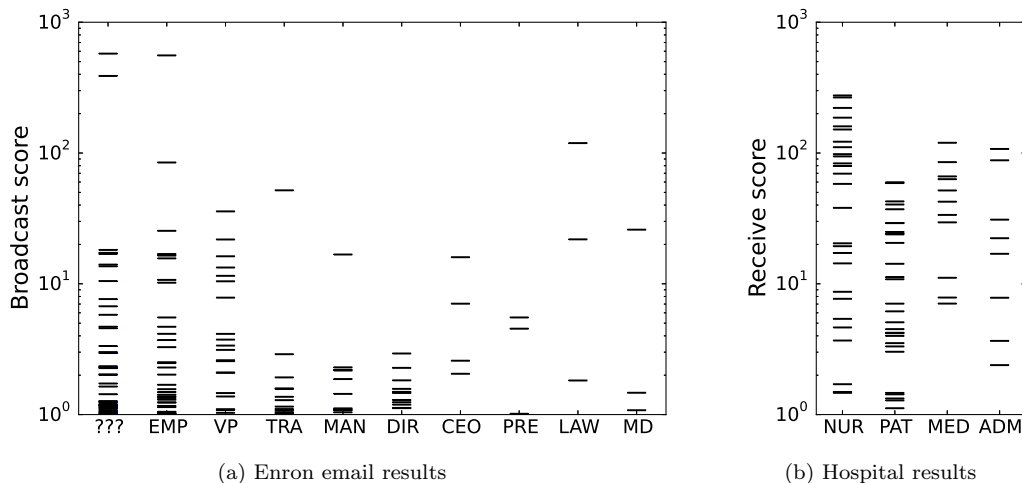
FIG. 3. Each marker (short horizontal line) represents an Enron employee or participant in the Sociopatterns experiment. The broadcast score, computed by Eq.(32), is displayed on the y-axis in (3a) and similarly the receive score in (3b) from Eq.(30) with $\alpha = 0.01$ and $\alpha = 0.005$ respectively. In both the employees (or participants) are divided into distinct categories shown along the x-axis. Abbreviations are given in Section IV.

V. RESULTS

A. Modeling

In Section III we derived formulae which predict the outcome of calculating the broadcast score for a large number of shuffled temporal edge-lists. The amount of error in these predictions is illustrated in Fig.(2) where we see that Eq.(18) gives accurate results regarding temporal edge lists with the time-column shuffled. The corresponding result, Eq.(27), appears to be less reliable however, owing to the computational cost of calculating the receive score multiple times, we chose only to test a very small sample. This contradicts the assumptions of the analytical model; particularly the assumption made in Section III B that the score $c_i(t)$ in an individual generation of the probabilistic model is well approximated by its mean, at time t, over many generations. It is likely that in a small data-set that there is a high variance in the distribution of receive scores and we expect the prediction to improve as the number of interactions increases. The creation of these "shortcut" formulae allowed us to perform data analysis on two large scale temporal edge-lists which would have otherwise taken an inconvenient amount of computation.

B. Data analysis

Using the method described in Section II we calculated the broadcast score for the Enron email corpus and the receive score for the Sociopatterns hospital ward experiment. We have chosen values of α that produce visually interesting figures; when too small the calculation of broadcast and receive scores are dominated by the con-

tribution from walks of length 1 and therefore become equivalent to the out-degree and in-degree respectively. Conversely, when α is too large, long walks dominate the scores and the edges with early timestamps determine the outcome. To anyone considering using these methods we recommend that a range of α values be tested, each one potentially exposing different information about the behavior of the system. For an in-depth analysis of α and its interpretation see [37].

The results are presented first in Fig.(3). In Fig.(4) we compare the result of each individual with their overall activity. We note two observations from Fig.(4): one Enron employee (a director) stands out as having an unusually high broadcast score when compared to a low amount of overall activity (broadcast rank 50, degree rank 125), and that patients in the hospital ward tend to have large receive scores considering their overall activity.

Fig.(5) shows the expected results of performing various shufflings, we can think of the y-axis in these plots as a measure of how much the score of each individual depends on temporal properties, and the x-axis for structural properties. We see that the outlier from the Enron data-set is, remarkably, unremarkable regarding both of these measures and neither alone can explain their high broadcast score (time-shuffled rank 86 and target-shuffled rank 104, both lower than the actual broadcast rank of 50). However, the fact that both shuffled ranks are higher than the degree rank suggests that the individual in question is sending emails very economically, i.e. sending to the most efficient recipients, and choosing the optimal moments in time to send. From this example it appears that the contributions from both factors add to the overall broadcast score.

An alternative interpretation is that the individual in question was feeding information into the network which was consequently being disseminated in a way that in-
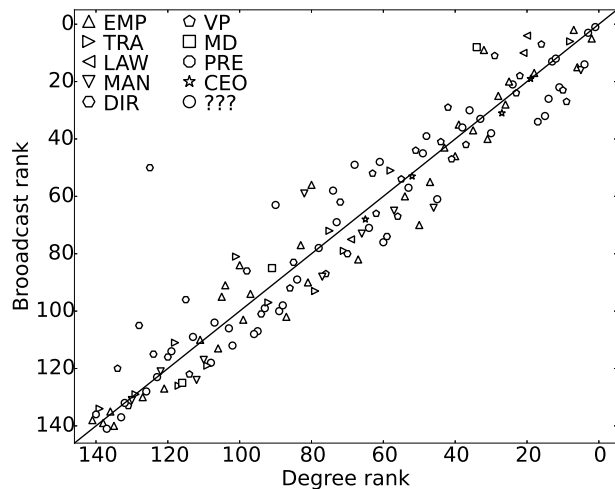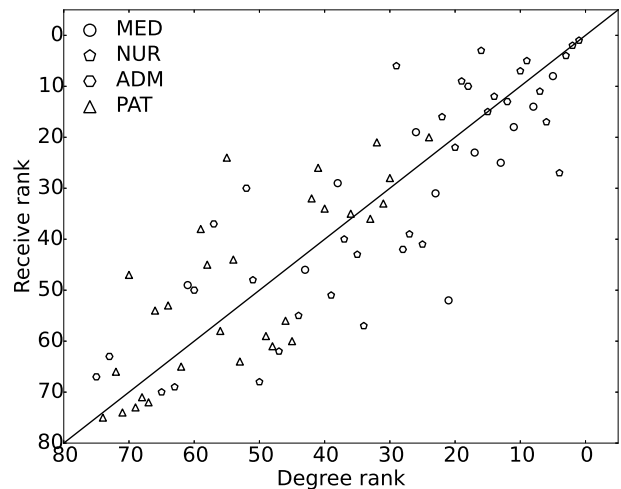
(a) Enron email directory   (b) Hospital contact network

FIG. 4. The rank according to broadcast score (left, computed by Eq.(32)) and receive score (right, computed by Eq.(30)), with $\alpha = 0.01$ and $\alpha = 0.005$ respectively, plotted against the out-degree (left) and in-degree (right). Each individual in the network is represented by a data point, their classification is given by their shape. The abbreviations in the legend are explained in Section IV. The one-to-one line is plotted as a visual aid to partition the nodes into two groups; those which have higher than expected scores (top left), and those who have lower than expected (bottom right).

flates their broadcast score (although similar results are not found for the CEOs who we would expect to be influential in the same way). The individual in question was a lobbyist for the corporation, after a very brief investigation we did not determine a particular reason why they should be significantly influential.

From Fig.(5b) it is apparent that shuffling the time column can cause large changes to the receive rank of a participant whereas the source-shuffling appears to be less effective. This is because the temporal activity of the participants deviates significantly from a Bernoulli process (that is assumed in the time-independent model). More specifically, nodes exist which are inactive towards the beginning of the sampling period but have a lot of activity at later time-steps. The receive score of these nodes is amplified by the exponential increase over time that is indicated by the very simple example in Section III D 1. Those which are active early on in the sampling period but have little or no activity at later times will have lower receive scores. When such effects dominate the outcome the effect of time-shuffling is significant.

While we do not discuss here the broadcast scores for the Hospital data, or the receive scores for the Enron email data, the results can be seen in Fig.(6), Fig.(7) and Fig.(8).

VI. DISCUSSION

As data-driven industries increasingly find value in targeting the most central, most influential, individuals, it is important to scrutinize the methods and tools that network science is promoting. The idea that there is one magic formula which can produce a meaningful result regardless of the system in question is firstly, wrong, and secondly, a counter-productive way of thinking. Here we have scrutinized the dynamic communicability metrics and found that temporal variation can have a stronger effect in some systems, like the hospital ward, than in others, like Enron. We have found efficient shortcut formulae to quantify the temporal component by randomizing the structural factors and likewise quantify the structural component by randomizing the temporal factors. Those who have data and wish to analyze dynamic communicability should use these methods to add more dimensions, and more depth, to their analysis.

When we look at the simple example of Fig.(1a), we can compute the broadcast scores and find that node A is ranked number one. We can then ask why A is the most influential broadcaster and find that it is not because it was the most active (C was in fact the most active), but because of a complex interplay of temporal and structural factors; A was the first to communicate, and importantly, one of those early edges was received by C who was subsequently the most active node. Looking at large data-sets it is tedious to try to deconstruct every sequence of contacts that caused each individual to achieve its score. Instead, we have introduced meaningful statistics, i.e. the results of shuffling, that provide insight into the interplay of temporal and structural factors.

Several specific phenomena are commonly found in social systems whose effects are nullified by the shuffling process. The structure (or topology) of complex networks has been extensively studied and many elements have repeatedly been found across different systems [38]. Degree heterogeneity is one such topological feature which
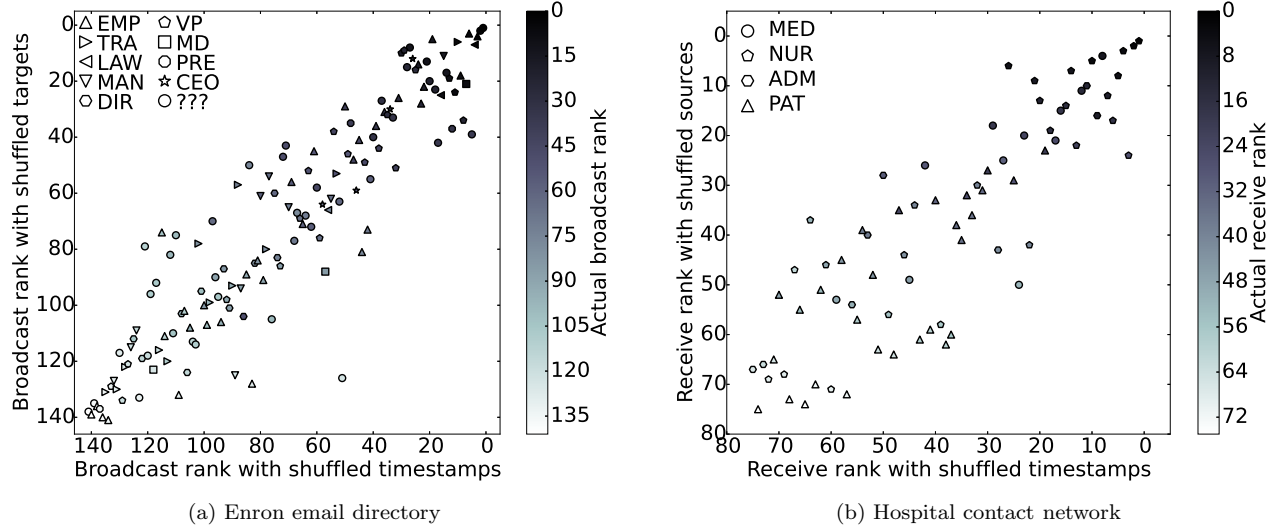
(a) Enron email directory
(b) Hospital contact network

FIG. 5. As demonstrated in Fig.(1c). On the y-axis we show the ranking of each node according to expectation of the broadcast score (left, computed using Eq.(28)) and receive score (right, computed with Eq.(27)) for the expected outcomes of the source (left) and target (right) shuffled networks (with $\alpha = 0.01$ and $\alpha = 0.005$ respectively). The x-axes show the expected scores for a time-shuffled network computed with Eqs.(17) and (18). The actual broadcast score computed with Eqs. (32) and (30) is shown by the darkness of the markers. Different roles are indicated by the marker shapes, the abbreviations are explained in Section IV. These results are also presented in a table in Appendix VIII.

is not nullified by shuffling. On the other hand, features like community structure, assortativity and clustering are likely to play a significant role in determining the communicability in most applications of this work [39–41]. Similarly, the non-shuffled data is likely to exhibit certain temporal features. Recent studies of communication data, similar to the Enron data-set, show that activity generally occurs in bursts [42]. Others focus on the effect of circadian cycles which are likely to occur in the hospital ward data [43].

Clustering and burstiness both increase the number of walks which revisit nodes. For most contagion processes these walks would not be permissible since, for example, many diseases can only be contracted once, similarly a piece of information can only be attained once (this is possibly the reason why bursty networks have been shown to slow the spread of information compared to their temporally shuffled equivalent [24]). This remains a fundamental problem of the dynamic communicability metric which should not be overlooked.

Another issue that ought to be considered when using the dynamic communicability metrics is the effect of a bounded sampling window. Take for example the simple example of Fig.(1a). Here A has the highest broadcast score because it is the first node to create outgoing edges. Had we observed the system just one time-step earlier we might have found one or more edges from C to A, thus making C the highest ranked broadcaster above A. This is a general issue; our analytical results tell us that the earlier interactions contribute exponentially more than those which occur later; therefore the first node involved in the first recorded interaction will, by chance, receive

an unduly high broadcast score. In the case of the receive score, interactions that occur late in the sample inflate the score of the involved nodes. The advancement of dynamic communicability presented in [28], that assumes infectiousness decays in the time between interactions, may mitigate these problems to some extent. We conclude this paper by suggesting two possible alternative solutions:

A. Control for temporal variation

Eq.(27) gives the expectation of the receive score based on temporal variation. It can therefore be considered as a control to compare to the actual score. Further, we suggest that a normalized version of the receive score would be a more appropriate measure to compare individuals in the same network. The normalized version is the ratio of the actual score, computed using Eq.(30), and its expectation, computed using Eq.(27).

B. Remove temporal variation

Alternatively, we ignore temporal variation altogether; in many circumstances this is sensible since the temporal variation over the duration of the sample is not usually expected to be the same in the future (unless perhaps it is driven by a cyclic process). Without knowledge of when each future interaction will occur, the Bernoulli process used in the time-independent model is a suitable choice. In such a case, the past data provides an esti-

mate of how active each node will be, but the timing of their interactions remains random. The matrix exponential in Eqs.(17) and (18), can be computed very efficiently to give these approximations to the receive score and broadcast score. Incidentally, the matrix exponential has previously been proposed as a centrality measure [25, 44].

[1] N. Perra and B. Gonçalves, in *Social Phenomena* (Springer International Publishing, 2015) pp. 59–83.
[2] L. Weng, F. Menczer, and Y.-Y. Ahn, Scientific reports **3**, 2522 (2013).
[3] S. Ronen, B. Gonalves, K. Z. Hu, A. Vespignani, S. Pinker, and C. A. Hidalgo, Proceedings of the National Academy of Sciences **111**, E5616 (2014), http://www.pnas.org/content/111/52/E5616.full.pdf.
[4] L. E. Quevillon, E. M. Hanks, S. Bansal, and D. P. Hughes, Scientific reports **5** (2015).
[5] G. Petri, P. Expert, F. Turkheimer, R. Carhart-Harris, D. Nutt, P. Hellyer, and F. Vaccarino, Journal of The Royal Society Interface **11**, 20140873 (2014).
[6] M. S. Granovetter, American journal of sociology , 1360 (1973).
[7] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai, Nature **411**, 41 (2001).
[8] A. E. Motter, A. P. de Moura, Y.-C. Lai, and P. Dasgupta, Physical Review E **65**, 065102 (2002).
[9] D. de Solla Price, Journal of the Association for Information Science and Technology **27**, 292 (1976).
[10] A.-L. Barabâsi, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, Physica A: Statistical mechanics and its applications **311**, 590 (2002).
[11] D. F. Gleich, SIAM Review **57**, 321 (2015), http://dx.doi.org/10.1137/140976649.
[12] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, V. Colizza, L. Isella, C. Régis, J.-F. Pinton, N. Khanafer, W. Van den Broeck, *et al.*, BMC medicine **9**, 87 (2011).
[13] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani, Proceedings of the National Academy of Sciences of the United States of America **103**, 2015 (2006).
[14] P. T. Johnson, J. C. De Roode, and A. Fenton, Science **349**, 1259504 (2015).
[15] R. Pastor-Satorras and A. Vespignani, Physical review letters **86**, 3200 (2001).
[16] C. Sanli and R. Lambiotte, Frontiers in Physics **3** (2015), 10.3389/fphy.2015.00079.
[17] S. Wang, T. Yang, Q. Zhang, and K. Zhao, arXiv preprint arXiv:1508.04417 (2015).
[18] C. Knappett, T. Evans, and R. Rivers, Antiquity **82**, 1009 (2008).
[19] P. Holme, The European Physical Journal B **88**, 1 (2015).
[20] T. O. Richardson and T. E. Gorochowski, Journal of The Royal Society Interface **12**, 20150705 (2015).
[21] P. Grindrod, M. C. Parsons, and E. Higham, Desmond J and, Physical Review E **83**, 046120 (2011).
[22] A. Masucci and G. Rodgers, Advances in Complex Systems **12**, 113 (2009).
[23] C. Sanli and R. Lambiotte, PLOS ONE **10**, e0131704 (2015).
[24] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki, Physical Review E **83**, 025102 (2011).
[25] E. Estrada and N. Hatano, Physical Review E **77**, 036111 (2008).
[26] E. Estrada, Physical Review E **88**, 042811 (2013).
[27] P. Grindrod and D. J. Higham, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **470**, 20130835 (2014).
[28] P. Grindrod and D. J. Higham, SIAM Review **55**, 118 (2013).
[29] T. Rogers, Journal of Complex Networks , cnu014 (2014).
[30] A. H. Al-Mohy and N. J. Higham, SIAM Journal on Matrix Analysis and Applications **31**, 970 (2009).
[31] M. Molloy and B. Reed, Random structures & algorithms **6**, 161 (1995).
[32] F. Chung and L. Lu, Annals of combinatorics **6**, 125 (2002).
[33] https://www.cs.cmu.edu/ ./enron/,.
[34] http://enrondata.org/assets/edo_enron-custodians data.html,.
[35] P. Vanhems, A. Barrat, C. Cattuto, J.-F. Pinton, N. Khanafer, C. Rgis, B.-a. Kim, B. Comte, and N. Voirin, PLoS ONE **8**, e73970 (2013).
[36] D. V. Greetham, Z. Stoyanov, and P. Grindrod, in *Computing and Combinatorics* (Springer, 2013) pp. 791–800.
[37] I. Chen, M. Benzi, H. H. Chang, and V. S. Hertzberg, arXiv preprint arXiv:1601.07586 (2016).
[38] M. E. Newman, SIAM review **45**, 167 (2003).
[39] A. Clauset, M. E. Newman, and C. Moore, Physical review E **70**, 066111 (2004).
[40] V. C. Barclay, T. Smieszek, J. He, G. Cao, J. J. Rainey, H. Gao, A. Uzicanin, and M. Salathé, PloS one **9**, e87042 (2014).
[41] N. Malik and P. J. Mucha, Chaos: An Interdisciplinary Journal of Nonlinear Science **23**, 043123 (2013).
[42] K.-I. Goh and A.-L. Barabási, EPL (Europhysics Letters) **81**, 48002 (2008).
[43] H.-H. Jo, M. Karsai, J. Kertész, and K. Kaski, New Journal of Physics **14**, 013055 (2012).
[44] M. Benzi and C. Klymko, Journal of Complex Networks **1**, 124 (2013).
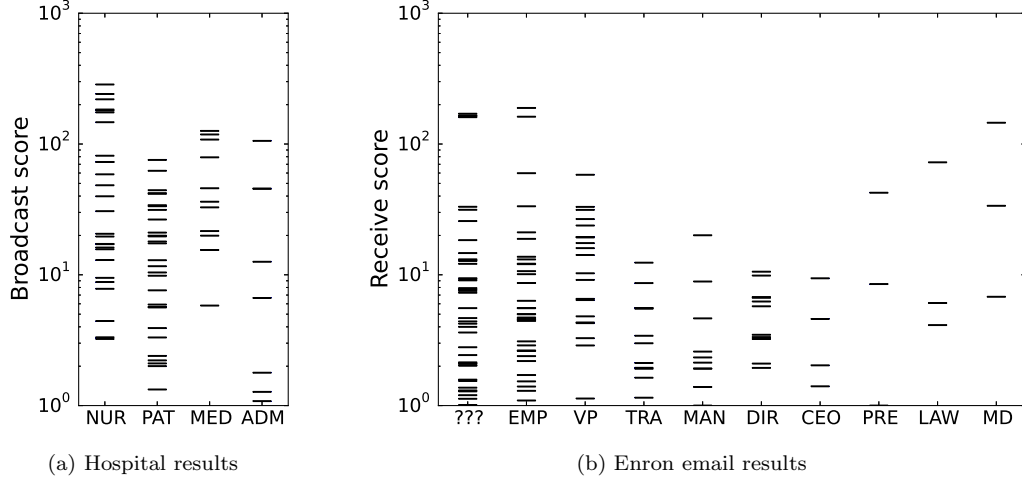
(a) Hospital results

(b) Enron email results

FIG. 6. Each marker (short horizontal line) represents an Enron employee or participant in the Sociopatterns experiment. The broadcast score, computed by Eq.(32), is displayed on the y-axis in (6a) and similarly the receive score in (6b) from Eq.(30) with $\alpha = 0.005$ and $\alpha = 0.01$ respectively. In both the employees (or participants) are divided into distinct categories shown along the x-axis. Abbreviations are given in Section IV.

VII. MODELING

A. Heterogeneous "send" and "receive" model

The Model:

In any given time-step, the probability that i has an out going edge is s_i, the probability that it has an incoming edge is r_i.

Making no further assumptions about who communicates with whom, letting $\mathbf{r}$ and $\mathbf{s}$ both be column vectors we have the general stochastic model with

$$P = \mathbf{s}\mathbf{r}^T = \begin{pmatrix} s_1 r_1 & s_1 r_2 & \cdots & s_1 r_N \\ s_2 r_1 & s_2 r_2 & \cdots & s_2 r_N \\ \vdots & \vdots & \ddots & \vdots \\ s_N r_1 & s_N r_2 & \cdots & s_N r_N \end{pmatrix}. \qquad (33)$$

There are at least two ways to find the expectation of broadcast and receive scores for this model. It is possible to write down an expression for the P^k which can then be substituted into Eq.(15). An alternative method is to solve Eq.(12) directly. First we express Eq.(12) in terms of our new variables:

$$\frac{\partial \hat{\mathbf{c}}(t)}{\partial t} = \alpha \hat{\mathbf{c}}(t)\mathbf{s}\mathbf{r}^T. \qquad (34)$$

Multiplying both sides on the right by s gives

$$\frac{\partial \hat{\mathbf{c}}(t)\mathbf{s}}{\partial t} = \alpha \hat{\mathbf{c}}(t)\mathbf{s}\mathbf{r}^T\mathbf{s}. \qquad (35)$$

which is a differential equation describing the time-evolution of $\hat{\mathbf{c}}(t)\mathbf{s}$, a scalar variable. This has the solution

$$\hat{\mathbf{c}}(t)\mathbf{s} = e^{\alpha[\mathbf{r}^T\mathbf{s}]t}. \qquad (36)$$

Substituting the result back into Eq.(34) we get

$$\frac{\partial \hat{\mathbf{c}}(t)}{\partial t} = \alpha \mathbf{r}^T e^{\alpha[\mathbf{r}^T\mathbf{s}]t}. \qquad (37)$$

Which has the solution

$$\hat{\mathbf{c}}(t) = \mathbf{1} + \left(\frac{e^{\alpha[\mathbf{r}^T\mathbf{s}]t} - 1}{\mathbf{r}^T\mathbf{s}} \right) \mathbf{r}^T. \qquad (38)$$

In a similar way one can show that the expectation of the broadcast score is

$$\hat{\mathbf{b}}(t) = \mathbf{1}^T + \left(\frac{e^{\alpha[\mathbf{r}^T\mathbf{s}]t} - 1}{\mathbf{r}^T\mathbf{s}} \right) \mathbf{s}. \qquad (39)$$

B. Time-dependent P matrix

The Model:

At time τ, the probability that i has an out going edge is s_i, the probability that it has an incoming edge is $r_i(\tau)$

Eq.(12) now becomes

$$\frac{\partial \hat{\mathbf{c}}(t)}{\partial t} = \alpha \hat{\mathbf{c}}(t)\mathbf{s}\mathbf{r}(t_0 + t)^T. \qquad (40)$$

Multiplying both sides on the right by $\mathbf{s}$ we get

$$\frac{\partial \hat{\mathbf{c}}(t)}{\partial t}\mathbf{s} = \alpha \hat{\mathbf{c}}(t)\mathbf{s}\mathbf{r}(t_0 + t)^T\mathbf{s}$$
$$\frac{\partial \hat{\mathbf{c}}(t)\mathbf{s}}{\partial t} = \alpha[\hat{\mathbf{c}}(t)\mathbf{s}][\mathbf{r}(t_0 + t)^T\mathbf{s}]. \qquad (41)$$

This equation now only includes scalar functions of t so we can solve to get

$$\hat{\mathbf{c}}(t)\mathbf{s} = \mathbf{1s}\exp\left(\alpha\int_0^t \mathbf{sr}(t_0+t')^T dt'\right). \qquad (42)$$

Substituting this back into Eq.(40) we have

$$\frac{\partial\hat{\mathbf{c}}(t)}{\partial t} = \alpha\mathbf{r}(t_0+t)^T\exp\left(\alpha\int_0^t \mathbf{sr}(t_0+t')^T dt'\right) \qquad (43)$$

Since $\mathbf{1}s = 1$.

C. Simple time-dependent example

The model:

At time τ_i person i is on the receiving end of r_i edges. As before, the number of outgoing edges is determined by a time-independent probability s_i.

Clearly, after N iterations the process will end so we use $t_0 = 0$ and $t_{\text{end}} = N$ as the initial and final conditions respectively. To find the broadcast score of a node i we solve Eq.(43) with

$$r_i(\tau) = r_i\delta(\tau - \tau_i) \qquad (44)$$

where r_i is a scalar and δ is the Dirac delta. The justification for this version of $r_i(\tau)$ is that the expected number of messages sent by i over some time-interval will be r_i if the time interval includes τ_i. Without loss of generality we can say $\tau_i = i$ meaning that node 1 sends first, then node 2 and so on. First we focus on expressing $\mathbf{s}\int_0^t \mathbf{r}(t_0+t')^T dt'$ in a simpler form. Since

$$\int_0^t r_j\delta(t'-j)dt' = \begin{cases} r_j & \text{if } j \leq t \\ 0 & \text{if } j \geq t \end{cases} \qquad (45)$$

(This result derives from the fact that the integral of the Dirac delta between $-\infty$ and t is the Heaviside step function $H(t)$.) we have

$$\mathbf{s}\int_0^t \mathbf{r}(t_0+t')^T dt' = \sum_{j=0}^N s_j\int_0^t r_j\delta(t'-j)dt'$$
$$= \sum_{j=0}^t r_j s_j. \qquad (46)$$

Substituting this into Eq.(43) then integrating over the whole sample gives

$$[\hat{c}_i(t')]_0^N = \int_0^N \alpha r_i\delta(t'-i)\exp\left(\alpha\sum_{j=0}^i r_j s_j\right)dt'. \qquad (47)$$

The integral is solved by the translation property of the Dirac delta and we have

$$\hat{c}_i(N) = 1 + \alpha r_i\exp\left(\alpha\sum_{j=0}^i r_j s_j\right). \qquad (48)$$

D. Incorporating empirical data

The model:

Let K_i be the set of edges for which i is the target node, and $\tau^{(k)}$ be the time at which edge k was present. As before, s_i is the time-independent probability for i to be the source an edge.

We achieve this by choosing

$$r_i(t) = \sum_{k\in K_i}\delta(t - \tau_i^{(k)}). \qquad (49)$$

We can choose the set K_i and the corresponding $\tau^{(k)}$ in a way that recreates exactly what is observed in the target and time columns of an empirical temporal edge-list. We introduce $R_i(\tau)$, the number of messages sent by i between time t_0 and time τ, this is expressed

$$R_i(\tau) = \int_{t_0}^\tau r_i(t')dt', \qquad (50)$$

giving

$$R_i(t_0+t) = \int_0^t r_i(t_0+t')dt', \qquad (51)$$

and therefore Eq.(43) can be expressed

$$\frac{\partial\hat{\mathbf{c}}(t)}{\partial t} = \alpha\mathbf{r}(t_0+t)^T e^{\alpha\sum_{j=1}^N s_j R_j(t_0+t)}. \qquad (52)$$

Integrating over the entire duration of the sample gives

$$[\hat{c}_i(t')]_0^{t_{\text{end}}-t_0}$$
$$= \alpha\int_0^{t_{\text{end}}-t_0}\left[\sum_{k\in K_i}\delta(t_0+t'-\tau^{(k)})\right]\exp\left(\alpha\sum_{j=1}^N s_j R_j(t_0+t')\right)dt'$$
$$= \alpha\sum_{k\in K_i}\int_{-\infty}^\infty \delta[t_0+t'-\tau^{(k)}]\exp\left(\alpha\sum_{j=1}^N s_j R_j(t_0+t')\right)dt'. \qquad (53)$$

Finally, using the translation property of the Dirac delta function we have

$$\hat{c}_i = 1 + \alpha\sum_{k\in K_i}\exp\left(\alpha\sum_{j\in N} s_j R_j(\tau_i^{(k)})\right). \qquad (54)$$
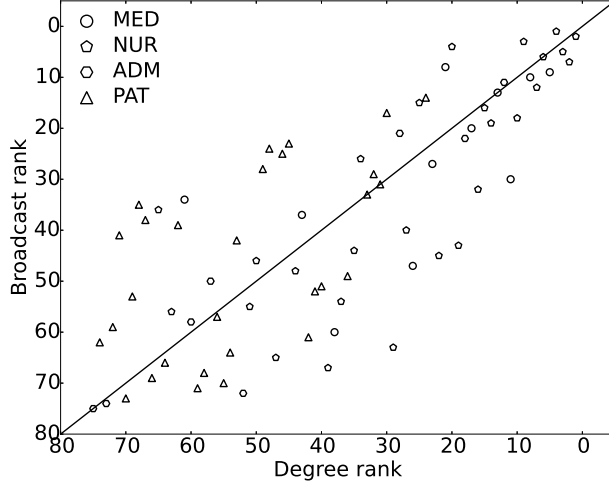
(a) Hospital contact network
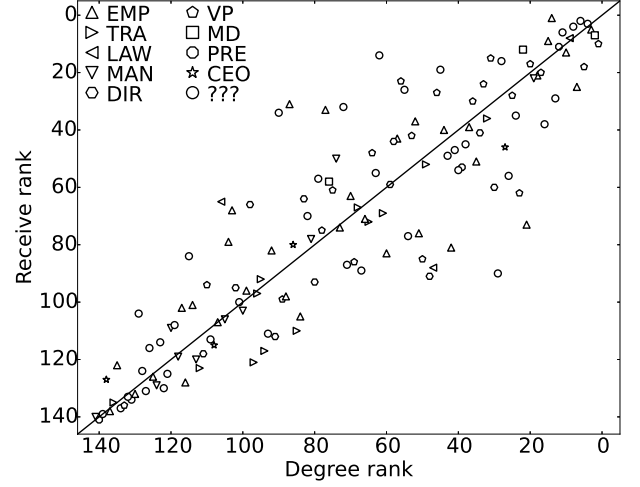
(b) Enron email directory

FIG. 7. The rank according to broadcast score (left, computed by Eq.(32)) and receive score (right, computed by Eq.(30)), with $\alpha = 0.005$ and $\alpha = 0.01$ respectively, plotted against the out-degree (left) and in-degree (right). Each individual in the network is represented by a data point, their classification is given by their shape. The abbreviations in the legend are explained in Section IV. The one-to-one line is plotted as a visual aid to partition the nodes into two groups; those which have higher than expected scores (top left), and those who have lower than expected (bottom right).
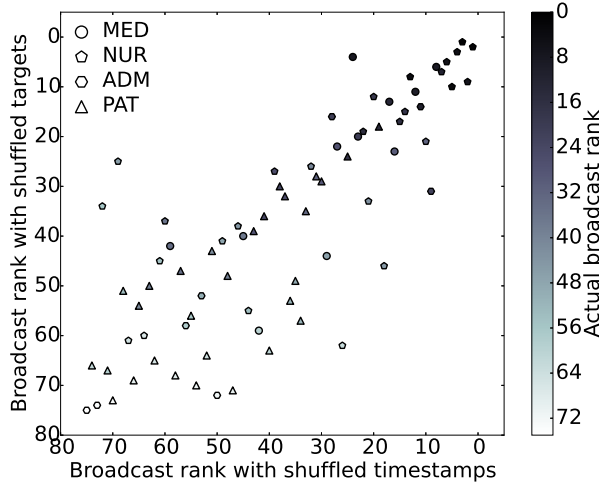
(a) Hospital contact network
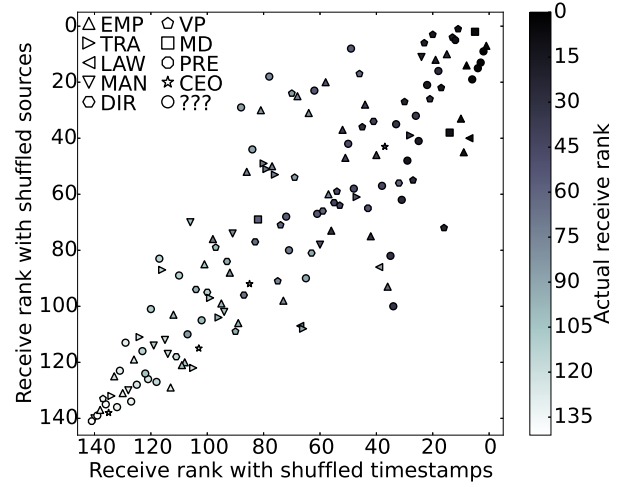
(b) Enron email directory

FIG. 8. As demonstrated in Fig.(1c). On the y-axis we show the ranking of each node according to expectation of the broadcast score (left, computed using Eq.(28)) and receive score (right, computed with Eq.(27)) for the expected outcomes of the source (left) and target (right) shuffled networks (with $\alpha = 0.005$ and $\alpha = 0.01$ respectively). The x-axes show the expected scores for a time-shuffled network computed with Eqs.(17) and (18). The actual broadcast score computed with Eqs. (32) and (30) is shown by the darkness of the markers. Different roles are indicated by the marker shapes, the abbreviations are explained in Section IV.

VIII.   RANKINGS

A.   Sociopatterns hospital ward receive-rank

Rank	None	Time-shuffled	Source-shuffled	Time and Source
1	1115 (NUR)	1115 (NUR)	1115 (NUR)	1115 (NUR)
2	1210 (NUR)	1210 (NUR)	1210 (NUR)	1210 (NUR)
3	1190 (NUR)	1207 (NUR)	1295 (NUR)	1295 (NUR)
4	1295 (NUR)	1295 (NUR)	1157 (MED)	1207 (NUR)
5	1109 (NUR)	1109 (NUR)	1190 (NUR)	1157 (MED)
6	1629 (NUR)	1164 (NUR)	1629 (NUR)	1164 (NUR)
7	1149 (NUR)	1193 (NUR)	1149 (NUR)	1193 (NUR)
8	1157 (MED)	1157 (MED)	1109 (NUR)	1144 (MED)
9	1205 (NUR)	1658 (ADM)	1205 (NUR)	1109 (NUR)
10	1658 (ADM)	1190 (NUR)	1098 (ADM)	1149 (NUR)
11	1193 (NUR)	1098 (ADM)	1144 (MED)	1221 (MED)
12	1196 (NUR)	1144 (MED)	1193 (NUR)	1098 (ADM)
13	1098 (ADM)	1114 (NUR)	1196 (NUR)	1159 (MED)
14	1144 (MED)	1149 (NUR)	1181 (NUR)	1196 (NUR)
15	1181 (NUR)	1181 (NUR)	1221 (MED)	1181 (NUR)
16	1625 (NUR)	1221 (MED)	1658 (ADM)	1190 (NUR)
17	1164 (NUR)	1159 (MED)	1164 (NUR)	1260 (MED)
18	1221 (MED)	1625 (NUR)	1130 (MED)	1658 (ADM)
19	1130 (MED)	1365 (PAT)	1625 (NUR)	1205 (NUR)
20	1365 (PAT)	1196 (NUR)	1260 (MED)	1114 (NUR)
21	1383 (PAT)	1205 (NUR)	1159 (MED)	1191 (MED)
22	1114 (NUR)	1245 (NUR)	1114 (NUR)	1625 (NUR)
23	1260 (MED)	1260 (MED)	1365 (PAT)	1148 (MED)
24	1547 (PAT)	1191 (MED)	1207 (NUR)	1365 (PAT)
25	1159 (MED)	1378 (PAT)	1148 (MED)	1245 (NUR)
26	1702 (PAT)	1629 (NUR)	1660 (MED)	1130 (MED)
27	1207 (NUR)	1148 (MED)	1383 (PAT)	1202 (NUR)
28	1378 (PAT)	1179 (ADM)	1671 (ADM)	1179 (ADM)
29	1660 (MED)	1130 (MED)	1378 (PAT)	1629 (NUR)
30	1671 (ADM)	1383 (PAT)	1202 (NUR)	1378 (PAT)
31	1148 (MED)	1352 (PAT)	1352 (PAT)	1352 (PAT)
32	1401 (PAT)	1202 (NUR)	1702 (PAT)	1383 (PAT)
33	1352 (PAT)	1391 (PAT)	1401 (PAT)	1391 (PAT)
34	1307 (PAT)	1702 (PAT)	1142 (NUR)	1105 (NUR)
35	1362 (PAT)	1362 (PAT)	1547 (PAT)	1108 (NUR)
36	1391 (PAT)	1307 (PAT)	1391 (PAT)	1362 (PAT)
37	1232 (ADM)	1374 (PAT)	1485 (NUR)	1142 (NUR)
38	1469 (PAT)	1393 (PAT)	1307 (PAT)	1660 (MED)
39	1202 (NUR)	1105 (NUR)	1469 (PAT)	1485 (NUR)
40	1142 (NUR)	1401 (PAT)	1232 (ADM)	1307 (PAT)
41	1245 (NUR)	1363 (PAT)	1362 (PAT)	1702 (PAT)
42	1179 (ADM)	1660 (MED)	1245 (NUR)	1401 (PAT)
43	1108 (NUR)	1395 (PAT)	1179 (ADM)	1168 (MED)
44	1701 (PAT)	1142 (NUR)	1108 (NUR)	1100 (NUR)
45	1460 (PAT)	1168 (MED)	1460 (PAT)	1393 (PAT)
46	1168 (MED)	1108 (NUR)	1261 (NUR)	1374 (PAT)
47	1784 (PAT)	1547 (PAT)	1613 (NUR)	1613 (NUR)
48	1261 (NUR)	1320 (PAT)	1701 (PAT)	1363 (PAT)
49	1152 (MED)	1100 (NUR)	1168 (MED)	1395 (PAT)
50	1209 (ADM)	1671 (ADM)	1191 (MED)	1246 (NUR)
51	1485 (NUR)	1327 (PAT)	1769 (PAT)	1261 (NUR)
52	1191 (MED)	1701 (PAT)	1784 (PAT)	1671 (ADM)
53	1769 (PAT)	1232 (ADM)	1152 (MED)	1327 (PAT)
54	1416 (PAT)	1469 (PAT)	1209 (ADM)	1701 (PAT)
55	1100 (NUR)	1385 (PAT)	1416 (PAT)	1547 (PAT)
56	1374 (PAT)	1209 (ADM)	1100 (NUR)	1385 (PAT)
57	1105 (NUR)	1399 (PAT)	1385 (PAT)	1232 (ADM)
58	1385 (PAT)	1460 (PAT)	1105 (NUR)	1460 (PAT)
59	1395 (PAT)	1152 (MED)	1363 (PAT)	1469 (PAT)
60	1393 (PAT)	1116 (NUR)	1374 (PAT)	1209 (ADM)
61	1363 (PAT)	1261 (NUR)	1395 (PAT)	1152 (MED)
62	1613 (NUR)	1769 (PAT)	1393 (PAT)	1320 (PAT)
63	1535 (ADM)	1377 (PAT)	1327 (PAT)	1238 (NUR)
64	1327 (PAT)	1485 (NUR)	1320 (PAT)	1769 (PAT)
65	1320 (PAT)	1323 (PAT)	1373 (PAT)	1116 (NUR)
66	1373 (PAT)	1416 (PAT)	1535 (ADM)	1416 (PAT)
67	1525 (ADM)	1613 (NUR)	1525 (ADM)	1377 (PAT)
68	1246 (NUR)	1305 (PAT)	1246 (NUR)	1399 (PAT)
69	1238 (NUR)	1246 (NUR)	1238 (NUR)	1305 (PAT)
70	1116 (NUR)	1784 (PAT)	1377 (PAT)	1784 (PAT)
71	1399 (PAT)	1373 (PAT)	1116 (NUR)	1323 (PAT)
72	1377 (PAT)	1238 (NUR)	1399 (PAT)	1373 (PAT)
73	1305 (PAT)	1535 (ADM)	1305 (PAT)	1535 (ADM)
74	1323 (PAT)	1332 (PAT)	1323 (PAT)	1332 (PAT)
75	1332 (PAT)	1525 (ADM)	1332 (PAT)	1525 (ADM)

B. Enron email broadcast rank

Rank	None	Time-shuffled	Target-shuffled	Time and Target
1	tana.jones (???)	tana.jones (???)	tana.jones (???)	tana.jones (???)
2	mark.taylor (EMP)	sara.shackleton (???)	sara.shackleton (???)	jeff.dasovich (EMP)
3	sara.shackleton (???)	mark.taylor (EMP)	jeff.dasovich (EMP)	sara.shackleton (???)
4	carol.clair (LAW)	carol.clair (LAW)	mark.taylor (EMP)	bill.williams (???)
5	jeff.dasovich (EMP)	marie.heard (???)	chris.germany (EMP)	mike.grigsby (MAN)
6	eric.bass (TRA)	jeff.dasovich (EMP)	eric.bass (TRA)	chris.germany (EMP)
7	steven.kean (VP)	mark.haedicke (MD)	carol.clair (LAW)	mark.taylor (EMP)
8	mark.haedicke (MD)	d..steffes (VP)	susan.scott (???)	eric.bass (TRA)
9	elizabeth.sager (EMP)	elizabeth.sager (EMP)	scott.neal (VP)	john.arnold (VP)
10	mary.hain (LAW)	eric.bass (TRA)	drew.fossum (VP)	scott.neal (VP)
11	richard.sanders (VP)	steven.kean (VP)	mike.grigsby (MAN)	phillip.love (???)
12	phillip.allen (???)	louise.kitchen (PRE)	david.delainey (CEO)	phillip.allen (???)
13	susan.scott (???)	richard.sanders (VP)	phillip.allen (???)	susan.scott (???)
14	bill.williams (???)	bill.williams (???)	sally.beck (EMP)	debra.perlingiere (???)
15	chris.germany (EMP)	mike.grigsby (MAN)	debra.perlingiere (???)	kimberly.watson (???)
16	mike.grigsby (MAN)	mary.hain (LAW)	john.arnold (VP)	steven.kean (VP)
17	sally.beck (EMP)	kim.ward (???)	bill.williams (???)	louise.kitchen (PRE)
18	drew.fossum (VP)	phillip.love (???)	elizabeth.sager (EMP)	sally.beck (EMP)
19	david.delainey (CEO)	chris.germany (EMP)	richard.sanders (VP)	david.delainey (CEO)
20	matthew.lenhart (EMP)	gerald.nemec (???)	gerald.nemec (???)	carol.clair (LAW)
21	gerald.nemec (???)	phillip.allen (???)	mark.haedicke (MD)	mary.hain (LAW)
22	phillip.love (???)	matthew.lenhart (EMP)	matthew.lenhart (EMP)	drew.fossum (VP)
23	scott.neal (VP)	kay.mann (EMP)	phillip.love (???)	d..steffes (VP)
24	d..steffes (VP)	sally.beck (EMP)	steven.kean (VP)	gerald.nemec (???)
25	kay.mann (EMP)	john.arnold (VP)	mary.hain (LAW)	matthew.lenhart (EMP)
26	debra.perlingiere (???)	david.delainey (CEO)	darron.giron (EMP)	darron.giron (EMP)
27	john.arnold (VP)	susan.scott (???)	mike.mcconnell (???)	john.lavorato (CEO)
28	darron.giron (EMP)	debra.perlingiere (???)	kay.mann (EMP)	kay.mann (EMP)
29	jane.tholt (VP)	scott.neal (VP)	kate.symes (EMP)	richard.sanders (VP)
30	mike.mcconnell (???)	drew.fossum (VP)	john.lavorato (CEO)	marie.heard (???)
31	john.lavorato (CEO)	darron.giron (EMP)	dan.hyvl (EMP)	kate.symes (EMP)
32	kimberly.watson (???)	barry.tycholiz (VP)	jane.tholt (VP)	elizabeth.sager (EMP)
33	lynn.blair (???)	kimberly.watson (???)	kimberly.watson (???)	lynn.blair (???)
34	louise.kitchen (PRE)	john.lavorato (CEO)	d..steffes (VP)	mark.haedicke (MD)
35	dan.hyvl (EMP)	jane.tholt (VP)	jeffrey.shankman (PRE)	errol.mclaughlin (EMP)
36	kim.ward (???)	dan.hyvl (EMP)	errol.mclaughlin (EMP)	mike.mcconnell (???)
37	errol.mclaughlin (EMP)	mike.mcconnell (???)	louise.kitchen (PRE)	kevin.presto (VP)
38	marie.heard (???)	kevin.presto (VP)	hunter.shively (VP)	kim.ward (???)
39	jeffrey.shankman (PRE)	errol.mclaughlin (EMP)	marie.heard (???)	dan.hyvl (EMP)
40	kate.symes (EMP)	lynn.blair (???)	lynn.blair (???)	michelle.lokay (EMP)
41	barry.tycholiz (VP)	michelle.cash (???)	michelle.lokay (EMP)	rod.hayslett (VP)
42	kevin.presto (VP)	kam.keiser (EMP)	kim.ward (???)	jane.tholt (VP)
43	tracy.geaccone (EMP)	rod.hayslett (VP)	rob.gay (???)	tracy.geaccone (EMP)
44	hunter.shively (VP)	stacy.dickson (EMP)	kevin.presto (VP)	barry.tycholiz (VP)
45	darrell.schoolcraft (???)	michelle.lokay (EMP)	chris.dorland (EMP)	mark.whitt (???)
46	michelle.lokay (EMP)	kenneth.lay (CEO)	fletcher.sturm (VP)	john.forney (MAN)
47	rod.hayslett (VP)	tracy.geaccone (EMP)	robin.rodrigue (???)	chris.dorland (EMP)
48	rob.gay (???)	jeffrey.shankman (PRE)	tracy.geaccone (EMP)	jeffrey.shankman (PRE)
49	robin.rodrigue (???)	fletcher.sturm (VP)	rod.hayslett (VP)	darrell.schoolcraft (???)
50	robert.badeer (DIR)	kate.symes (EMP)	andrea.ring (???)	kam.keiser (EMP)
51	tori.kuykendall (TRA)	susan.bailey (???)	barry.tycholiz (VP)	hunter.shively (VP)
52	greg.whalley (VP)	mark.whitt (???)	greg.whalley (VP)	kenneth.lay (CEO)
53	kenneth.lay (CEO)	tori.kuykendall (TRA)	tori.kuykendall (TRA)	bill.rapp (???)
54	fletcher.sturm (VP)	hunter.shively (VP)	john.forney (MAN)	lindy.donoho (EMP)
55	chris.dorland (EMP)	martin.cuilla (MAN)	michelle.cash (???)	fletcher.sturm (VP)
56	peter.keavey (EMP)	james.derrick (LAW)	peter.keavey (EMP)	shelley.corman (VP)
57	bill.rapp (???)	jeffrey.hodge (MD)	mark.guzman (TRA)	martin.cuilla (MAN)
58	michelle.cash (???)	jeff.skilling (CEO)	darrell.schoolcraft (???)	tori.kuykendall (TRA)
59	daren.farmer (MAN)	andy.zipper (VP)	kenneth.lay (CEO)	kevin.hyatt (DIR)
60	lindy.donoho (EMP)	darrell.schoolcraft (???)	larry.may (DIR)	andrea.ring (???)
61	mark.whitt (???)	chris.dorland (EMP)	daren.farmer (MAN)	rob.gay (???)
62	larry.may (DIR)	bill.rapp (???)	martin.cuilla (MAN)	andy.zipper (VP)
63	benjamin.rogers (???)	greg.whalley (VP)	mark.whitt (???)	greg.whalley (VP)
64	john.forney (MAN)	dutch.quigley (???)	jeff.skilling (CEO)	dutch.quigley (???)
65	martin.cuilla (MAN)	lindy.donoho (EMP)	rick.buy (MAN)	jeff.skilling (CEO)
66	andy.zipper (VP)	shelley.corman (VP)	james.derrick (LAW)	rick.buy (MAN)
67	shelley.corman (VP)	patrice.mims (???)	patrice.mims (???)	t..lucci (EMP)
68	jeff.skilling (CEO)	monique.sanchez (???)	dutch.quigley (???)	robin.rodrigue (???)
69	monique.sanchez (???)	peter.keavey (EMP)	shelley.corman (VP)	james.derrick (LAW)
70	kam.keiser (EMP)	rick.buy (MAN)	benjamin.rogers (???)	jonathan.mckay (DIR)
71	dutch.quigley (???)	rob.gay (???)	lindy.donoho (EMP)	jim.schwieger (TRA)
72	mark.guzman (TRA)	robin.rodrigue (???)	bill.rapp (???)	larry.may (DIR)
73	rick.buy (MAN)	thomas.martin (VP)	kam.keiser (EMP)	monique.sanchez (???)
74	kevin.hyatt (DIR)	kevin.hyatt (DIR)	mike.carson (EMP)	michelle.cash (???)
75	james.derrick (LAW)	larry.may (DIR)	dana.davis (???)	mark.guzman (TRA)
76	andrea.ring (???)	joe.parks (???)	andy.zipper (VP)	thomas.martin (VP)
77	stacy.dickson (EMP)	john.forney (MAN)	monique.sanchez (???)	teb.lokey (MAN)
78	patrice.mims (???)	jim.schwieger (TRA)	kevin.ruscitti (TRA)	patrice.mims (???)
79	jim.schwieger (TRA)	john.zufferli (EMP)	judy.hernandez (???)	diana.scholtes (TRA)
80	jonathan.mckay (DIR)	daren.farmer (MAN)	jim.schwieger (TRA)	peter.keavey (EMP)

Rank	None	Time-shuffled	Target-shuffled	Time and Target
81	kevin.ruscitti (TRA)	t..lucci (EMP)	stacy.dickson (EMP)	john.zufferli (EMP)
82	t..lucci (EMP)	jonathan.mckay (DIR)	larry.campbell (???)	daren.farmer (MAN)
83	sandra.brawner (DIR)	richard.ring (EMP)	kevin.hyatt (DIR)	stacy.dickson (EMP)
84	geir.solberg (EMP)	andrea.ring (???)	t..lucci (EMP)	sandra.brawner (DIR)
85	jeffrey.hodge (MD)	judy.townsend (EMP)	jonathan.mckay (DIR)	matt.smith (???)
86	geoff.storey (DIR)	robert.badeer (DIR)	thomas.martin (VP)	danny.mccarty (VP)
87	thomas.martin (VP)	teb.lokey (MAN)	sandra.brawner (DIR)	cara.semperger (EMP)
88	teb.lokey (MAN)	mark.guzman (TRA)	jeffrey.hodge (MD)	larry.campbell (???)
89	matt.smith (???)	doug.gilbert-smith (MAN)	judy.townsend (EMP)	dana.davis (???)
90	john.zufferli (EMP)	diana.scholtes (TRA)	matt.smith (???)	benjamin.rogers (???)
91	judy.townsend (EMP)	geoff.storey (DIR)	john.zufferli (EMP)	jeffrey.hodge (MD)
92	danny.mccarty (VP)	danny.mccarty (VP)	jason.williams (???)	ryan.slinger (TRA)
93	diana.scholtes (TRA)	sandra.brawner (DIR)	diana.scholtes (TRA)	joe.parks (???)
94	jay.reitmeyer (EMP)	jay.reitmeyer (EMP)	teb.lokey (MAN)	sean.crandall (DIR)
95	holden.salisbury (EMP)	charles.weldon (???)	sean.crandall (DIR)	jason.williams (???)
96	frank.ermis (DIR)	matt.smith (???)	paul.thomas (???)	paul.thomas (???)
97	ryan.slinger (TRA)	benjamin.rogers (???)	charles.weldon (???)	jay.reitmeyer (EMP)
98	larry.campbell (???)	ryan.slinger (TRA)	danny.mccarty (VP)	geoff.storey (DIR)
99	joe.parks (???)	cara.semperger (EMP)	ryan.slinger (TRA)	mike.carson (EMP)
100	dana.davis (???)	geir.solberg (EMP)	geir.solberg (EMP)	geir.solberg (EMP)
101	sean.crandall (DIR)	sean.crandall (DIR)	geoff.storey (DIR)	kevin.ruscitti (TRA)
102	cara.semperger (EMP)	kevin.ruscitti (TRA)	susan.pereira (EMP)	judy.hernandez (???)
103	mike.carson (EMP)	jason.wolfe (???)	frank.ermis (DIR)	charles.weldon (???)
104	paul.y'barbo (???)	scott.hendrickson (???)	robert.badeer (DIR)	judy.townsend (EMP)
105	andrew.lewis (DIR)	holden.salisbury (EMP)	joe.parks (???)	holden.salisbury (EMP)
106	charles.weldon (???)	keith.holst (DIR)	jay.reitmeyer (EMP)	theresa.staab (EMP)
107	jason.williams (???)	susan.pereira (EMP)	cara.semperger (EMP)	paul.y'barbo (???)
108	paul.thomas (???)	frank.ermis (DIR)	holden.salisbury (EMP)	vladi.pimenov (???)
109	jason.wolfe (???)	albert.meyers (EMP)	jeff.king (MAN)	don.baughman (TRA)
110	susan.pereira (EMP)	dana.davis (???)	paul.y'barbo (???)	jeff.king (MAN)
111	mike.swerzbin (TRA)	paul.y'barbo (???)	theresa.staab (EMP)	susan.pereira (EMP)
112	judy.hernandez (???)	larry.campbell (???)	andrew.lewis (DIR)	doug.gilbert-smith (MAN)
113	theresa.staab (EMP)	mike.swerzbin (TRA)	scott.hendrickson (???)	jason.wolfe (???)
114	scott.hendrickson (???)	theresa.staab (EMP)	jason.wolfe (???)	harry.arora (VP)
115	mike.maggi (DIR)	mike.carson (EMP)	vince.kaminski (MAN)	frank.ermis (DIR)
116	keith.holst (DIR)	don.baughman (TRA)	don.baughman (TRA)	john.griffith (MD)
117	jeff.king (MAN)	jason.williams (???)	tom.donohoe (???)	eric.saibi (TRA)
118	vladi.pimenov (???)	john.griffith (MD)	vladi.pimenov (???)	mike.swerzbin (TRA)
119	don.baughman (TRA)	paul.thomas (???)	mike.maggi (DIR)	scott.hendrickson (???)
120	richard.shapiro (VP)	vladi.pimenov (???)	mike.swerzbin (TRA)	keith.holst (DIR)
121	vince.kaminski (MAN)	judy.hernandez (???)	harry.arora (VP)	richard.ring (EMP)
122	harry.arora (VP)	mike.maggi (DIR)	eric.saibi (TRA)	vince.kaminski (MAN)
123	susan.bailey (???)	pam.butler (???)	john.griffith (MD)	susan.bailey (???)
124	doug.gilbert-smith (MAN)	jeff.king (MAN)	keith.holst (DIR)	mike.maggi (DIR)
125	john.griffith (MD)	andrew.lewis (DIR)	doug.gilbert-smith (MAN)	robert.badeer (DIR)
126	eric.saibi (TRA)	vince.kaminski (MAN)	susan.bailey (???)	tom.donohoe (???)
127	richard.ring (EMP)	harry.arora (VP)	cooper.richey (MAN)	albert.meyers (EMP)
128	tom.donohoe (???)	eric.saibi (TRA)	richard.ring (EMP)	clint.dean (TRA)
129	clint.dean (TRA)	richard.shapiro (VP)	joe.stepenovitch (VP)	andrew.lewis (DIR)
130	albert.meyers (EMP)	tom.donohoe (???)	clint.dean (TRA)	cooper.richey (MAN)
131	cooper.richey (MAN)	clint.dean (TRA)	joe.quenet (TRA)	joe.stepenovitch (VP)
132	pam.butler (???)	cooper.richey (MAN)	albert.meyers (EMP)	pam.butler (???)
133	joe.stepenovitch (VP)	joe.stepenovitch (VP)	pam.butler (???)	steven.merris (???)
134	joe.quenet (TRA)	stephanie.panus (EMP)	richard.shapiro (VP)	richard.shapiro (VP)
135	stephanie.panus (EMP)	joe.quenet (TRA)	steven.merris (???)	monika.causholli (EMP)
136	stanley.horton (PRE)	brad.mckay (EMP)	phillip.platter (EMP)	mark.fisher (???)
137	steven.merris (???)	stanley.horton (PRE)	stanley.horton (PRE)	phillip.platter (EMP)
138	brad.mckay (EMP)	phillip.platter (EMP)	mark.fisher (???)	stephanie.panus (EMP)
139	phillip.platter (EMP)	steven.merris (???)	monika.causholli (EMP)	stanley.horton (PRE)
140	monika.causholli (EMP)	monika.causholli (EMP)	brad.mckay (EMP)	joe.quenet (TRA)
141	mark.fisher (???)	mark.fisher (???)	stephanie.panus (EMP)	brad.mckay (EMP)