TemporalPrestige

Nicholas Winsley

2025-03-19

Contents

Introduction

Epidemiology is a subject of much contemporary relevance. Quarantine and case isolation have been shown to be effective policies to counteract pandemics (Auranen et al. 2023). With the emergence of reliable contact tracing technology, prophylactic identification and isolation of high-risk individuals could be practical in the future. This study investigates methods for the identification of high-risk individuals in a temporal network.

We will begin by defining what is meant by "high-risk" in this context. Government's wish to isolate individuals who are most instrumental in the spread of a disease. Therefore, risk is a combination of the likelihood of a particular individual becoming contagious due to prior

contacts, and the likelihood of the individual spreading the illness via future contacts. In practice, future contacts are not known, therefore this study will assume perfect contact information up to the current time, and no future contact information.

The sociosphere of an individual can be modeled as a series of contacts, each paired with a measure of proximity. Each contact is written as (i, j, t), where i and j are the interacting individuals, and t is the time of the interaction. Note that (i, j, t) is semantically no different from (j, i, t). We abstractly represent each individual as a node, and each contact event by an undirected edge. This representation is called a temporal network. The Susceptible-Infected model, abbreviated as SI, is an epidemiological framework in which susceptible individuals can become infected due to a contact with an infectious individual. In the SI paradigm, we assume that infected individuals stay infected indefinitely. Social Network Analysis (SNA) is the study of individuals and their relationships with others. SNA involves constructing a network to model a real-world situation of interest, and calculating metrics related to the network structure. These metrics can be broadly categorized into two types: Population-level measures and Individual-level measures. This study is primarily concerned with individual-level measures. Centrality is a commonly-used individual-level measure which is defined as the influence of a particular individual within a network. Prestige (sometimes called status or rank) is a similar individual-level measure for directional graphs, which only considers incoming dyads.

To date, research on the SI model has largely been focused on computational simulations of many epidemics on a temporal network. In this paper, I propose several measures of prestige in temporal networks, and I apply them to the SI epidemiology framework.

Methods

The New Zealand Ministry of Health (MoH) comissioned a pilot study of the "CovidCard", A portable device which used bluetooth technology to record contacts between carriers of the device. We use this data to validate existing centrality metrics, and develop novel alternatives.

Study Site

The CovidCard pilot study was carried out in Ngontotahā, a small settlement situated in the Rotorua district. Ngontotahā was chosen because it met several key criteria, namely compactness, geographical isolation, small population size and high sociodemographic diversity.

Sampling

Adults 19 years of age or older who live in Ngontotahā West and East were recruited to participate in a seven-day study. Additionally, people who live outside these boundaries but work within the Ngontotahā Village were also permitted to take part in the trial. In total, 1,191 people participated in the study. At the end of the trial period, a subset of 158 participants from the main trial were contacted by MoH case investigators to establish contacts that they had over the trial period using a modified version of the MoH case investigation protocol. The data collected through these investigations was important, as it was meant to provide a "ground truth" against which the CovidCard data could be validated.

Data Description

The CovidCard is a bluetooth device developed for detecting close-contact events between carriers. Each card advertises it's presence and detects signals from other cards. Algorithms evaluate the radio signal strength indicator (RSSI) of close-contacts in real-time, and the signal strength is aggregated over 15 minute time intervals. Each interval was classified as either < 1 meter, < 2 meter and < 4 meter proximity, and the total number of intervals belonging to each class was summed over a two-hour period. The cards can hold up to 128 contact events in short-term cache memory at any given time, of which some are recorded in long-term flash memory. An interaction was recorded in flash memory if it was longer than 2 minutes in duration, and the RSSI exceeded -62dBm (roughly corresponding to a distance of less than 4 meters).

Data Preparation

The last day of the trial period saw an anomalously high number of close-contacts, most likely because participants congregated at a single location to return the cards. For this reason, contact events which occurred on the last day of the trial were omitted. Data which could not be cross-verified by case investigation was removed, and duplicate contact events were omitted.

Temporal Centrality

Generalization of conventional centrality measures to temporal networks requires a high-granularity representation of the network as a graph. Kempe, Kleinberg, and Kumar (2000) proposed a graph where the edge weights are contact times, however this model fails to account for differential rates of transmission. Kim and Anderson (2012) proposed a more general solution. Without loss of generality, consider a set of contact times, $t = (1, 2, 3, \cdots, T)$. We can construct a time-ordered graph, G, with vertices V and edges E, where for all $v_t, u_k \in V$, an edge $e(v_t, u_k) \in E$ can only exist if k = t + 1. We can construct this graph for any temporal network without loss of information. In practice, computational constraints may require aggregation of contact times and thus loss of information.

To illustrate this idea, we will use a simple example with five individuals. The contact data is shown in the table below:

First Individual	Second Individual	Time of contact
a	b	1
b	d	2
a	c	2
d	e	1

Figure 1 shows the corresponding representation as a directed graph.

Define the distance of the temporal shortest path length over time interval [i, j], denoted by $d_{i,j}(v, u)$, as the smallest $d = j - n$, where $i \leq n \leq j$ and there is a path from v_n to u_j. Thus, in Figure 1, the shortest path distance $d_{1,3}(a, b)$ is two, with the temporal shortest path being a_1-b_2-b_3. By representing a temporal network as a high-granularity digraph, we
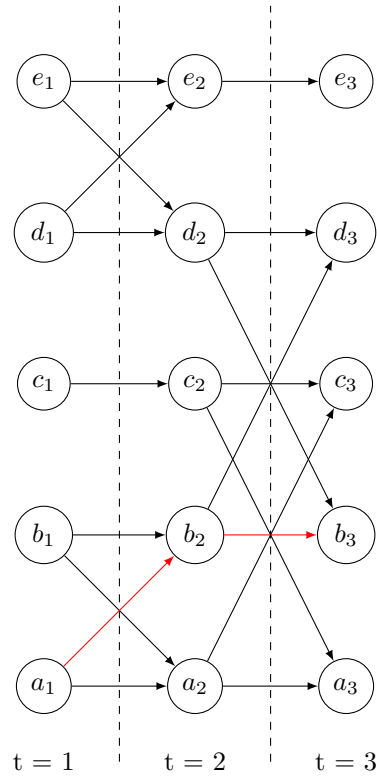
4

Figure 1: A simple temporal network represented as a digraph. The temporal shortest path from a to b is shown in red.

can now generalize conventional measures of prestige and centrality to temporal networks.

Temporal Closeness and Proximity Prestige

In this study, I am primarily concerned with the likelihood of infection, which is closely related to the more general notion of prestige. In a directional network, a prestigious node is the object of many ties i.e. has many incoming connections. This is a distinct concept from centrality, which is also concerned with outgoing connections. Many conventional measures of centrality are ill-defined in directional graphs, owing to the fact that directional graphs are not necessarily strongly-connected. Due to this limitation, we usually only consider nodes in the influence domain of node i i.e. the set of all nodes from whom n_i is reachable. Lin (1976) proposed the following measure of prestige for directional relations, called the proximity prestige:

$$P_p(n_i) = \frac{I_i/(g-1)}{\sum d(n_j, n_i)/I_i} \tag{1.}$$

where I_i is the size of the influence domain of node i, and g is the total number of nodes in the network. For temporal networks, I propose a modified version

$$P_p(u_i) = \frac{I_u/(g-1)}{\sum_{v \in N} d_{0,i}(v, u)/I_u} \tag{2.}$$

where N is the influence domain of node u at time i. Intuitively, this is the proportion of the network covered by the influence domain, divided by the average temporal distance over the influence domain. When the influence domain is empty, the proximity prestige is defined to be 0. Kim and Anderson [2012] proposed temporal closeness, a similar metric which considers all time intervals $[t, i], t \in [0, i-1]$.

$$C_{i,j} = \sum_{i \leq t < j} \sum_{u \in V} \frac{1}{d_{t,j}(v, u)} \tag{3.}$$

When u is unreachable from v over [t, j], $d_{t,j}(v, u) = \infty$. We cover cases where the denominator is infinite by assuming that $\frac{1}{\infty} = 0$. Note that as we are considering a directional graph,

$d_{t,j}(v, u)$ is not equivalent to $d_{t,j}(u, v)$. To turn (3.) into a prestige measure, we simply reverse the direction of the paths to get:

$$C_{i,j}^{P} = \sum_{i \leq t < j} \sum_{u \in V} \frac{1}{d_{t,j}(u, v)} \tag{4.}$$

We will call this the temporal prestige. The temporal prestige can be normalized by dividing by $(|V| - 1)(j - i)$.

Kim and Anderson (2012) showed that by considering all time-intervals, temporal closeness improves the quality of estimates in some situations. By this same token, we can modify equation (2.) to obtain

$$P_p(u_i) = \sum_{t=0}^{i-1} \frac{I_{t,i,u}/(g-1)}{\sum_{v \in N} d_{t,i}(v, u)/I_{t,i,u}} \tag{5.}$$

where $I_{t,i,u}$ is the influence domain of u over the time interval [t, i]. We will call this the temporal proximity prestige. The temporal proximity prestige can be normalized by dividing by i.

When each edge is associated with a probability of transmission, as is the case with epidemiology models, the probability of a path may be of greater interest than the temporal length. In this case, we can generalize existing methods by considering a digraph where the edge weights are the natural log of the probability. Figure 2 shows a graph of this kind.

In accordance with the Susceptible-Infected (SI) framework, the probability of transmission from an individual to them self is assumed to be one, and hence the natural log becomes 0. This representation has several useful mathematical properties. Consider a path, P, starting at v_i and ending at u_j. The probability of this path is equal to $\prod_{k=i}^{j} E_k$, where E is the list of transmission probabilities of path P. The probability of path P can be calculated by:

$$e^{\sum_{w \in E} log(w)}$$

In order to generalize conventional measures of centrality, we need to define distance in terms
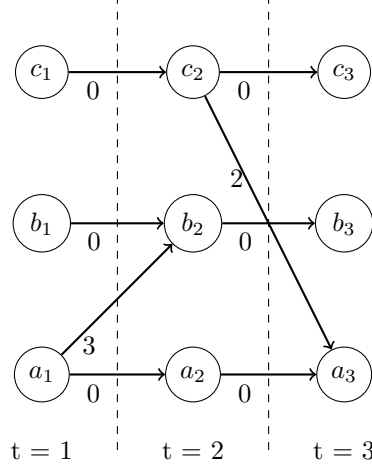
Figure 2: A simple temporal network represented as a digraph

of probability. To this end, I propose a new measure of distance, which I call the stochastic distance:

$$d^{\mathbf{P}} = 1 - \sum_{w \in E} log(w)$$

Note that the stochastic distance ranges from 1 to ∞, therefore we can adjust equations (4.) and (5.) to get

$$C_{i,j}^{P} = \sum_{i \leq t < j} \sum_{u \in V} \frac{1}{d_{t,j}^{\mathbf{P}}(u, \upsilon)} \tag{4.}$$

and

$$P_p(u_i) = \sum_{t=0}^{i-1} \frac{I_{t,i,u}/(g-1)}{\sum_{\upsilon \in N} d_{t,i}^{\mathbf{P}}(\upsilon, u)/I_{t,i,u}} \tag{5.}$$

where $d_{t,i}^{\mathbf{P}}(\upsilon, u)$ is the shortest stochastic path distance between υ and u over the time interval $[t, i]$. Note that the shortest stochastic path is not necessarily equal to the shortest temporal path, and thus we cannot use conventional algorithms for calculating geodesics. Nevertheless, it can be shown that for the representation in Figure 2, all stochastic shortest paths to u_i can be calculated in $O(i|V|^2)$ time using a modified version of the Reversed Evolution Network (abbreviated as REN) algorithm proposed by Hanke and Foraita (2017). Algorithm 1 shows

8

the psuedocode for this algorithm.

Result: Temporal Closeness Prestige
Input : Data file with each row representing a contact and a probability of transmission
Output: Temporal Prestige for a given node

contacts <- List of contact times sorted in increasing order. Each contact time is a data
  structure with a map of nodes to out-neighbours.
while *Data file has next line* **do**
  Read line
  Add line to contacts using bisection search
end
tcp <- Temporal closeness prestige
reachable <- Set of reachable nodes
Add target node to reachable
sums <- Map of node id to log of the highest probability path (obtained by summing
  edge weights)
back <- Pointer to final contact time in contacts list
while *back is not null* **do**
  temp <- Map of updated path lengths for this iteration
  foreach *node in reachable* **do**
    out-neighbours <- back.out-neighbours[node]
    foreach *neighbour in out-neighbours* **do**
      Add neighbour to reachable set
      weight <- Edge weight of connection between node and neighbour
      temp[neighbour] = Max(temp[neighbour], sums[node] + weight)
    end
  end
  foreach *node in reachable* **do**
    sums[node] = Max(temp[node], sums[node])
    tcp += $\frac{1}{1-sums[node]}$
  end
  Decrement back pointer
end
Return tcp

Algorithm 1: Modified version of the REN algorithm.

Absorption Prestige

Rocha and Masuda (2014) proposed TempoRank, an extension of the illustrious Google
PageRank algorithm to temporal networks. TempoRank considers the stationary distribution
of a random walk through the temporal network. However, TempoRank does not generalize

well to epidemiology modelling, where infection may be a permanent state. Intuitively, some sort of aggregation over all previous contacts would seem to be a preferable option. In general, I propose two broad categories of temporal centrality:

- Centrality of an individual at a particular point in time
- Centrality of an individual aggregated over all time points

Temporal proximity prestige and temporal closeness prestige belong to the latter category. For the former category, I recommend TempoRank or the more general in-weight rank (which will be discussed later). In this section, I propose an aggregate metric for temporal networks. Without loss of generality, we will consider a temporal network consisting of a set of nodes, N, and positive integer contact times, $t = (1, 2, 3, \cdots, T)$. Let $p_{ijk}(t)$ denote the probability of transmission for the k'th contact between individual i and individual j, at time t. Let $n_{ij}(t)$ denote the number of contacts between i and j at time t. Define the transition probability matrix for each contact time as:

$$
\mathbf{B}_{ij}(t) = \begin{cases} 1 & i = j, s_i(t) = 0 \\ 0 & i \neq j, s_{ij}(t) = 0 \\ \prod_{m \in N, m \neq i} \prod_{k=1}^{n_{im}(t)} (1 - p_{imk}(t)) & i = j, s_i(t) > 0 \\ (1 - \mathbf{B}_{ii})(s_{ij}(t)/s_i(t)) & i \neq j, s_{ij}(t) > 0 \end{cases} \tag{6.}
$$

where $s_{ij}(t)$ and $s_i(t)$ are defined as:

$$
s_{ij}(t) = 1 - \prod_{k=1}^{n_{ij}(t)} (1 - p_{ijk}(t))
$$

$$
s_i(t) = \sum_{j \in N, j \neq i} s_{ij}(t)
$$

Denote by $\mathbf{B}_i(t)$ the transition matrix obtained by taking $\mathbf{B}(t)$, and setting all entries in the i'th row to zero, except the diagonal entry (which is necessarily one). The walk $\mathbf{B}_i = (\mathbf{B}_i(1), \mathbf{B}_i(2), \cdots, \mathbf{B}_i(T))$ is an absorbing random walk, and the product $C_i^A(t) = \mathbf{B}_i(1)\mathbf{B}_i(2)\cdots\mathbf{B}_i(t)$ will be called the temporal absorption prestige for individual i, at time t.

Example

A simple example network of five nodes is shown in Example_Network.csv. For simplicity, a constant transmission probability was used for all contact events. In total, 500 simulations were run on this network. The absorption centrality and the number of times infected were calculated for the final contact time. Table 1 shows the correlation between the absorption centrality and the probability of infection by the end of the simulation.

Table 1: Pearson correlation between temporal absorption centrality and infected proportion over 500 simulations

	Correlation
p = 0.2	0.6614322
p = 0.1	0.5206837

In-weight Rank

Rocha and Masuda (2014) proposed the use of in-weights as a measure of prestige. However, the in-weights fail to account for the prestige of the sender. To address this limitation, I propose a novel approach, which I will call in-weight rank. Recall the transition matrix, $\mathbf{B}(t)$, from equation 6. Denote by $P(t)$ the prestige vector at time t. I propose a recursive formula for the prestige at time t

$$P(t) = P(t-1)\mathbf{B}(t)$$

In other words, the prestige of node i at time t is

$$P_i(t) = \sum_{j=1}^{N} P_j(t-1)B_{ij}(t)$$

Define the initial prestige at time 1 as the solution to the equation

$$P(1)^T = \mathbf{B}(1)^T P(1)^T \tag{7.}$$

This is similar to the method proposed by Seeley (1949). Using this method, the initial rank vector is the eigenvector corresponding to an eigenvalue of 1. Since $\mathbf{B}$ is a transition matrix, the columns of $\mathbf{B}^T$ all sum to 1, and thus the matrix has an eigenvalue of 1. Note, however, that this does not demonstrate uniqueness of the initial rank vector. By taking the transpose on both sides, it is easy to see that the initial rank vector is the stationary distribution of the transition matrix. A unique stationary distribution does not exist in general. For uniqueness, the following conditions are sufficient:

- For any two nodes i and j, i is reachable from j and vice-versa
- At least one diagonal element is > 0

For the transition matrix $\mathbf{B}(t)$, the second condition necessarily holds. For sparse networks, the first condition is unlikely to hold, in which case I propose several rudimentary solutions:

- Use a known prior distribution for the starting ranks.
- Aggregate transition matrices (by matrix multiplication) until the condition holds.
- Use the stationary distribution of the entire temporal network as the initial rank.
- Arbitrarily choose a solution to equation 7.
- If the state of the network is known at time 0, set the rank to 1 for infected individuals, and 0 for healthy individuals.

In general, the initial rank may be estimable from exogenous prior information (e.g. disease tests, qualitative case investigation etc). It should be noted that the rank can only be meaningfully compared between nodes in the same network, at the same time point. For cross-comparison between networks and/or time points, the ranks must be scaled accordingly. If the initial rank is chosen to be the stationary distribution of the T-step transition matrix (where T is the final time of the temporal network), then the formula reduces and the in-weight rank is identical to the TempoRank.

Simulation

Simulation is an effective method for ascertaining properties of a temporal network. Some classical models have been deterministic; the differential equation model of Kermack and McKendrick (1927) is a notable example. However, deterministic models typically rely on simplifying assumptions, and thus do not capture the full granularity of the network. Stochastic simulations typically follow the standard Markovian framework, in which we assume transmission depends only on the current state of the network. The contact times can be thought of as discrete snapshots of the network, and it is assumed that during each snapshot, only one "hop" can occur. In other words, if we have two contacts (i, k, t) and (k, j, t), i cannot infect j (via k) at time t. The standard Markovian framework typically employs simplifying assumptions to ensure ease of implementation. The Susceptible-Infected-Recovered (SIR) model is a common implementation which assumes that all individuals are susceptible at the beginning, and once infected they remain infectious for a recovery period. Once recovered, individuals cannot be infected again. Other common implementations include the Susceptible-Infected (SI) model and the Susceptible-Infected-Susceptible (SIS) model.

Algorithms

For simulation, we use the event-based algorithm first proposed by Kiss et al. (2017), and described in great detail by Holme (2021). In this algorithm, contacts are conveniently stored an adjacency list format (each node is represented by a data structure with a list of neighbours). For each neighbour, a sorted list of contact times (and associated transmission probabilities) is stored. All infection events are stored in a min-heap ordered by infection time. At each step, the earliest infection is removed and processes. This continues until no infection events remain in the heap. When node i is infected, we iterate through the list of neighbours and add the earliest time of infection (if any) to the heap. If we assume a constant transmission probability, the index of the first infectious contact follows a geometric distribution. We can sample this index by

$$\lceil \frac{log(1-X)}{log(1-\beta)} \rceil$$

where β is the fixed transmission probability and $X \sim Uniform(0, 1)$.

Constant Transmission Probability

The constant transmission probability assumption is commonplace in the literature, most likely because it greatly speeds up the simulation algorithm. To test this assumption, we compare the observed probability of infection for 751 individuals for two different algorithms (run on the same temporal network). Probabilities were calculated using an arbitrarily defined logistic regression model. In the first algorithm, the probability of transmission is calculated separately for every contact. In contrast, the second algorithm assumes that all contacts between any pair of nodes (i, j) have the same probability (taken to be the average over all contacts between i and j). In total, 1000 simulations were run for each algorithm. For each run, the indicator variable, I_i, is 1 if node i was infected, and 0 otherwise. The observed probability of infection for node i is the average of I_i over all 1000 runs.

Denote by $\hat{P}^{\mathbf{1}}$ the vector of observed probabilities for algorithm 1 (and likewise for algorithm 2). Thus, $\hat{P}_i^{\mathbf{1}} - \hat{P}_i^{\mathbf{2}}$ is the observed difference between the two algorithms for node i. We wish to test whether this difference is equal to 0. The test statistic is

$$ Z = \frac{\hat{P}_i^{\mathbf{1}} - \hat{P}_i^{\mathbf{2}}}{\hat{P}_i(1 - \hat{P}_i)(\frac{1}{1000} + \frac{1}{1000})} $$

where $\hat{P}_i$ is the pooled proportion under the null hypothesis ($P_i^{\mathbf{1}} = P_i^{\mathbf{2}}$). This is a well-known test statistic which follows a standard normal distribution if the proportions are equal. In order to test for equality of the vectors $P^{\mathbf{1}}$ and $P^{\mathbf{2}}$, we must adjust the significance level of each individual test to correct for multiple comparisons. The Bonferroni correction, proposed by Dunn (1961), is an adjustment which controls the family-wise Type I error rate. For a given significance level α, the Bonferroni correction guarantees a family-wise Type I error rate which is $\leq \alpha$. It does this by setting the test-wise significance level to $\frac{\alpha}{N}$, where N is the number of tests. The Bonferroni correction is ideal for this experiment because it makes no assumptions about independence between the individual tests. Note that the Bonferroni correction is conservative, meaning it lacks power for rejecting the null hypothesis.

Holm's method, proposed by Holm (1979), is a more powerful alternative to the Bonferroni correction. Holm's method tests the hypotheses iteratively, updating the p-value at each step. First, we sort the list of p-values in increasing order, and we begin the sequential significance tests from the lowest p-value. The algorithm starts with a significance level of $\frac{\alpha}{N}$. If the first result is non-significant, we test the second result with a significance level of $\frac{\alpha}{N-1}$. In general, the i'th test statistic is tested with a significance level of $\frac{\alpha}{N-i+1}$. Holm's method also guarantees a family-wise Type I error rate of $\leq \alpha$, however it offers greater power than the Bonferroni correction.

The Šidák correction, proposed by Šidák (1967), is a slightly less conservative alternative to the Bonferroni correction in which we set the test-wise significance level to $1 - (1-\alpha)^{\frac{1}{N}}$. This is the exact solution when tests are independent, conservative when tests are positively correlated and liberal when tests are negatively correlated.

References

Dunn, Olive Jean. 1961. "Multiple Comparisons Among Means." *Journal of the American Statistical Association* 56 (293): 52–64.

Hanke, Moritz, and Ronja Foraita. 2017. "Clone Temporal Centrality Measures for Incomplete Sequences of Graph Snapshots." *BMC Bioinformatics* 18: 1–18.

Holm, Sture. 1979. "A Simple Sequentially Rejective Multiple Test Procedure." *Scandinavian Journal of Statistics*, 65–70.

Holme, Petter. 2021. "Fast and Principled Simulations of the SIR Model on Temporal Networks." *Plos One* 16 (2): e0246961.

Kempe, David, Jon Kleinberg, and Amit Kumar. 2000. "Connectivity and Inference Problems for Temporal Networks." In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing*, 504–13.

Kermack, William Ogilvy, and Anderson G McKendrick. 1927. "A Contribution to the Mathematical Theory of Epidemics." *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 115 (772): 700–721.

Kim, Hyoungshick, and Ross Anderson. 2012. "Temporal Node Centrality in Complex

Networks." *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 85 (2): 026107.

Kiss, István Z, Joel C Miller, Péter L Simon, et al. 2017. "Mathematics of Epidemics on Networks." *Cham: Springer* 598 (2017): 31.

Lin, Nan. 1976. "Foundations of Social Research." *(No Title).*

Rocha, Luis EC, and Naoki Masuda. 2014. "Random Walk Centrality for Temporal Networks." *New Journal of Physics* 16 (6): 063023.

Seeley, John R. 1949. "The Net of Reciprocal Influence. A Problem in Treating Sociometric Data." *Canadian Journal of Experimental Psychology* 3: 234.

Šidák, Zbyněk. 1967. "Rectangular Confidence Regions for the Means of Multivariate Normal Distributions." *Journal of the American Statistical Association* 62 (318): 626–33.