

**Revisiting the Small-World Phenomenon: Efficiency Variation and Classification of Small-World Networks**

**Tore Opsahl**

Independent researcher

**Antoine Vernet**

Imperial College Business School  
Imperial College London

**Tufool Alnuaimi**

Imperial College Business School  
Imperial College London

**Gerard George**

Lee Kong Chian School of Business  
Singapore Management University

Accepted for publication *Organizational Research Methods*

September 2016

Opsahl, T., Vernet, A., Alnuaimi, T., George, G. 2017. Revisiting the Small-World Phenomenon: Efficiency Variation and Classification of Small-World Networks. *Organizational Research Methods*, in press.

## ABSTRACT

Research has explored how embeddedness in small-world networks influences individual and firm outcomes. We show that there remains significant heterogeneity among networks classified as small-world networks. We develop measures of the efficiency of a network, which allow us to refine predictions associated with small-world networks. A network is classified as a small-world network if it exhibits a distance between nodes that is comparable to the distance found in random networks of similar sizes—with ties randomly allocated among nodes—in addition to containing dense clusters. To assess how efficient a network is, there are two questions worth asking: (i) ‘what is a compelling random network for baseline levels of distance and clustering?’ and (ii) ‘how proximal should an observed value be to the baseline to be deemed comparable?’. Our framework tests properties of networks, using simulation, to further classify small-world networks according to their efficiency. Our results suggest that small-world networks exhibit significant variation in efficiency. We explore implications for the field of management and organization.

Tore Opsahl (tore@opsahl.co.uk) is a Senior Quant Research Associate in the Global Markets Risk Analytics group at Bank of America. His work includes developing novel credit and liquidity risk models as well as incorporating contagion effects in existing ones.

Antoine Vernet (a.vernet@imperial.ac.uk) is a post-doctoral research associate in the management department at Imperial College Business School, and a member of the Business Analytics Centre with KPMG. His research is focused on the influence of one’s network of relationships in activity, such as gathering resource, identifying opportunities and choosing collaborators.

Tufool Alnuaimi (t.alnuaimi@imperial.ac.uk) is an Assistant Professor of Data Science and Innovation at Imperial College Business School. Her current research deploys analytics in workplace to understand the factors that improve team performance in different settings.

Gerry George (ggeorge@smu.edu.sg) is Dean and Lee Kong Chian Chair Professor of Innovation and Entrepreneurship at Lee Kong Chian School of Business at Singapore Management University. His recent work examines strategic and organizational factors, such as business models and organizational design, and their impact on innovation and performance.

## ACKNOWLEDGEMENT

The authors want to thank Filip Agneessens, Oliver Alexy, Jana Diesner, Bernie Hogan, Renaud Lambiotte, Matthieu Latapy, José Javier Ramasco, Felix Reed-Tsochas, Ammon Salter, John Skvoretz, Alessandro Vespignani, and Anne ter Wal as well as workshop participants at Imperial College London (UK), the ISI institute (Turin, Italy), the Sunbelt Social Networks Conference 2011 and 2013 (St. Pete Beach, Florida, USA and Hamburg, Germany). Gerry George gratefully acknowledges the support of The Lee Foundation, Singapore.

## Introduction

The small-world network phenomenon has captured the imagination of scholars since the Milgram studies ([Milgram, 1967](#)). Milgram and his colleagues found that individuals in Nebraska were typically six steps away from a stockbroker in Boston. While containing weaknesses ([Watts, 2003](#)), the “six-degrees-of-separation” experiment is often cited as evidence that we live in a small-world network. This property of networks has stimulated research in a number of fields.

The concept of a small-world network went through several iterations, from early formulations in the 1950s and 1960s ([Milgram, 1967](#); [Rapoport & Horvath, 1961](#)) to a formalized model by [Watts and Strogatz \(1998\)](#). The perception behind the small-world network concept was the seemingly counterintuitive notion that individuals are more closely knit together than they think. In other words, most individuals are indirectly connected to each other, and there are fewer intermediaries between them than one would expect. The graph-theoretic definition by [Watts and Strogatz \(1998\)](#) relies on two different properties of networks: path-length and clustering. Path-length is a measure of the number of intermediaries between two individuals in the network. Clustering measures the local density of ties around a specific individual; in other words, the proportion of an individual’s contacts which are directly linked to each other. The combination of short path-length and clustering means that, in a small-world network, individuals are—on average—a short distance from everyone else in the network, while at the same time retaining dense local connections. The groundbreaking advance of the model designed by Watts and Strogatz is their use of random networks and regular lattice as baseline to establish whether or not a network exhibits small-world network properties. Random networks are networks in which ties are allocated at random, usually within a set of constraints ([Erdos & Renyi, 1959](#)). Lattices are perfectly regular networks in which every node is linked to a set number of its neighbors

([Wasserman & Faust, 1994](#)). [Watts and Strogatz \(1998\)](#) paved the way for researchers to test the presence or absence of small-world network properties in their networks.

[Watts and Strogatz \(1998\)](#) observed that most “real” networks exhibit small-world properties. This captured the interest of researchers across disciplines. In organizational research, small-world networks have been deemed to exist readily and have a multiplicity of outcomes. They have been found to be an efficient way of organizing information flow between firms ([Verspagen & Duysters, 2004](#)) and to enhance innovative performance of firms embedded in such networks ([Schilling & Phelps, 2007](#)). In the case of board interlocks, small-world network properties have also been shown to be resilient to turnover among members of the network ([Davis, Yoo, & Baker, 2003](#)). At the individual level, small-world networks have been shown to enhance performance of creative teams on Broadway shows, both in terms of revenues and critical appeal ([Uzzi & Spiro, 2005](#)). Small-world networks have also been linked with higher creativity and faster diffusion of innovation in networks of inventors ([Fleming & Marx, 2006](#)). In a community of software developers, small-world networks have been shown to enhance programmers’ productivity ([Singh, 2010](#)). However, not all results have been consistent. In their review of the literature, Uzzi and colleagues have found that effects on performance of network members are generally positive, but with varying patterns across studies ([Uzzi, Amaral, & Reed-Tsochas, 2007](#)). At the firm level, Fleming, King and Juda (2007) failed to find evidence for any benefits of small-world networks on firms’ innovation within geographical regions. However, they found a positive effect of one of the components of small-world networks—path-length—on firms’ innovation. This suggest that the results at the firm level might not be robust, or that they are dependent on other conditions.

The fact that results vary across studies raises a number of questions. Perhaps small-world network characteristics have a complex relationship (neither linear nor quadratic) with different

measures of performance? Is our current measurement apparatus crude and imprecise, unable to capture the likely behavior of a network? These are the questions at the core of this paper. In fact, we are hardly the first to notice that research on small-world networks is far from complete: management scholars have recognized the existence of weaknesses in the current apparatus used to assess whether a network is a small-world network ([Steen, Macaulay, & Kastle, 2011](#); [Uzzi \*et al.\*, 2007](#)). However, those studies focused on limitations of the link between the measure employed and the mechanisms it was said to capture, on the interpretation of the results, or on the limitations of the data used. Measurement issues and their solutions have not featured highly in past research on small-world networks—this is what we tackle, as we believe there are ways to enhance our measurement apparatus. There are three limitations of the current measurement method that need to be overcome: (i) the widely used formula to calculate small-world networks is sensitive to large differences between the results for path-length and those for clustering; (ii) the current apparatus is only defined for the simplest types of networks, and applying it to more complex networks might have unintended consequences on the result; and (iii) there is a fundamental ambiguity regarding what makes a random network appropriate for comparison.

We build on the small-world network literature to design measures of efficiency of networks that refine our understanding of small-world network characteristics and their behavior, and overcome the three limitations highlighted above. By efficiency, we refer to the degree to which the structure of a network will facilitate circulation of flux. In order to do so, we draw on advances in network measurement. Especially, we will measure how observed networks compare to random networks in terms of clustering and path-length through simulation. High clustering and short path-length, compared to the ones obtained with comparable random networks, increase the speed of diffusion of information through a network. Our framework refines prediction about

networks by dividing the class of small-world networks into subclasses of more or less efficient networks. We find many networks that are characterized as small-world networks do not exhibit efficient properties, which suggests that there is substantial variation in the properties of small-world networks. Thus, we offer a rationale for some of the inconsistent results in the literature, and tools that will allow researchers to better assess the characteristics of the networks, with direct implications for social science and organization research.

### **A Framework to estimate network efficiency**

The small-world network phenomenon opens new avenues for management and organizational scholarship. From this perspective, it is more than just firms' and individuals' characteristics or network positions that matter. The structure of the relationships between actors embedded in the network—firms or individuals—helps to predict the outcomes. The literature has identified several mechanisms that explain why small-world networks are beneficial in different contexts, such as firm alliances or creative teams ([Schilling & Phelps, 2007](#); [Uzzi & Spiro, 2005](#)). Those mechanisms usually postulate that the benefits of being embedded in a small-world network are related to the speed of information diffusion coupled with the existence of dense clusters, believed to quicken the execution of new ideas. The speed of diffusion is usually attributed to the short average path-length (the number of links one need to traverse to join any two individuals in the network) found in small-world networks. Dense clusters are subgroups of individuals or firms that share many relations within the group; in other words, they are cohesive subgroups. Small-world networks lead to increased efficiency because the network retains cohesive groups that are more likely to mobilize quickly and effectively to bring a new idea to fruition ([Fleming \*et al.\*, 2007](#); [Fleming & Marx, 2006](#); [Schilling & Phelps, 2007](#))<sup>i</sup>, even while favoring diffusion of knowledge.

Those two mechanisms—speed of information diffusion and ability to mobilize—rely on the two properties of small-world networks defined in Watts and Strogatz’s model ([Watts & Strogatz, 1998](#)): path-length and clustering coefficient. Their model starts from a regular ring lattice network, in which each node is connected to its  $k$ -nearest neighbors, and progressively rewires the edges between nodes to transform this network into a random network ( $k$  is a parameter equal to how many of its neighbors every node is connected to). Regular lattices have high clustering by definition because a node and its neighbors are densely connected locally. Random networks have short average path-length because there is no segregation between groups of nodes, due to the randomness of the ties. Figure 1 illustrates that the transition between a ring lattice to a random network depends on the proportion of ties that are rewired (ties for which the receiving node is randomly swapped for a new node); the first panel of Figure 1 shows a ring lattice, the second panel shows a small-world network, and the third a random network. Watts and Strogatz observe that there is a transition phase between the regular lattice and the random network, during which the network retains clustering levels close to those of a regular lattice while already displaying path-length similar to the ones found in random networks. They go on to show that this, in turn, affects the speed at which a disease will spread in a network. Specifically, they find that diseases spread at a similar speed in both small-world networks and random networks, but at a much slower speed on a regular ring lattice.

---Insert Figure 1 about here---

The work of Watts and Strogatz generated significant interest in many fields, including social sciences and management. However, they offer little guidance as to how to use the model to estimate small-world network properties for networks that are not generated by their model. The guidance they offer is summarized in the legend of Table 1 of their paper:  $L_{actual} \geq L_{random}$  but

$C_{actual} \gg C_{random}$ , where  $L$  is the average path-length and  $C$  the clustering for a given network, while  $L_{random}$  and  $C_{random}$  are the average path-length and clustering for an equivalent random network. The emerging practice among social scientists to estimate small-world network properties was mirrored on their model: measuring path-length and clustering, and comparing them to those found in a comparable random network. The comparison takes the form of a ratio of ratios:

$$\frac{CC_o/CC_r}{PL_o/PL_r}$$

(where  $CC_o$  and  $PL_o$  are the clustering coefficient and the path-length in the observed network and  $CC_r$  and  $PL_r$  are those in the comparable random network). This has proved to work well when evaluating whether or not networks generated with the Watts and Strogatz model displayed small-world network properties.

However, this method is not in the original paper by Watts and Strogatz, and has potential caveats. First, if the two ratios that compose the final ratio are very different, the largest ratio is going to have an inflated influence on the final statistic. Second, this method is appropriate for undirected and unweighted networks; however, many network observed by management scholars are directed or weighted. For example, records of transactions between exchange partners are directed: there is a buyer and a seller. Furthermore, those are also weighted, as transactions vary in amount. As a consequence, extending the method for directed and weighted networks seems like a natural step. Defining measures for directed or weighted networks offers researchers an alternative to the current common practice of simplifying their network (by ignoring weights or direction of ties) in order to use the framework, which—as our results will show—has unintended consequences on the results obtained. Thirdly, the term random network covers a wide family of synthetic networks, and there are many different subclasses of random networks. It begs the



question: which type of random network should be used for comparison? While it seems there is no consensus in the current literature, we offer guidelines for researchers to choose the family of random networks best suited for their purposes.

The simplest random networks, often referred to as Erdos-Renyi or Bernoulli networks, are constructed by using the same number of nodes as those in the network under study, and assigning each dyad a uniform probability of having a tie based on the observed density (Erdos & Renyi, 1959). In these networks, the network distance has been shown to grow with the logarithm of the number of nodes ([Chung & Lu, 2002](#)). Baseline values based on random networks often rely on formula using the numbers of nodes and ties as input (Erdos & Renyi, 1959; [Molloy & Reed, 1995](#); [Opsahl, Colizza, Panzarasa, & Ramasco, 2008](#)). The main advantage of this approach is that network properties can often be approximated using equations with the number of nodes and ties only. For example, the network distance is approximately the ratio between the logarithm of nodes and the logarithm of the average degree of nodes ([Newman, 2001b](#); [Watts & Strogatz, 1998](#)). The disadvantage of this approach is that the distribution of ties across nodes is uniform or Poisson (Erdos & Renyi, 1959): such distributions rarely exists in reality.

Most real-world distributions are skewed ([Albert, Jeong, & Barabási, 1999](#); [Molloy & Reed, 1995](#)), implying that some nodes are hubs, which should reduce network distance ([Albert et al., 1999](#); [Dodds, Muhamad, & Watts, 2003](#); [Granovetter, 2003](#); [Newman, 2001b](#); [Simşek & Jensen, 2008](#); [Yeates & Beeby, 2006](#)). In other words, the skewed (scale free) nature of degree distributions in most observed networks leads one to question whether the values for comparison should be obtained from a random network with a similar degree distribution, rather than uniform ones. Indeed, some scholars have made the case that this is the best way to obtain baseline values ([Uzzi & Spiro, 2005](#)).

We refine the estimation of both path-length and clustering. Our framework is an extension of the method used to estimate small-world properties that allow for the evaluation of how efficient a network is. We conceptualize efficiency as the combination of characteristics that lead to a high speed of information diffusion (e.g. short path-length) and ease of resource mobilization (e.g. clustering). To determine the efficiency of a network, we measure how much the network under study differs statistically from the same characteristics in comparable random networks, for each of the aforementioned characteristics. Unlike previous works—that have relied on equations—we create ensembles of random networks using randomization procedures. While costly in computation, simulations lead to more accurate baseline values and expose the distribution of network distances in similar networks. If network distances found in random networks follow a normal or Gaussian distribution, the standard deviation can be calculated. In our analysis, the distributions of values for random networks were normally distributed for all networks. However, it is an additional feature of our framework that the analyst can evaluate, by looking at the distribution of values, when conclusions about a network might be less reliable due to non-normality of the distribution.

Our framework differs from the previous method on several accounts. The usual practice to estimate small-world network properties relies on the values obtained from one random network and compares them to the network under study. The limitation of this approach is that there is no way for the analyst to know if the values obtained from the random network are representative of the population of random networks generated with the same properties. For example, for Erdos-Renyi networks, this population consists of all random networks with the same number of nodes and ties. We overcome this limitation by drawing several random networks from that family, allowing us to observe the distribution of values in that population. This approach is similar in

spirit to several methods developed to overcome challenges associated with working with network data. For example, multiple regression quadratic assignment procedure (MRQAP) overcomes the limitations of standard significance tests in the presence of non-independence between observations (as is the case in a network), by performing multiple permutations on the matrix of ties and attributes between actors in the network ([Dekker, Krackhardt, & Snijders, 2007](#)). Exponential random graph models aim to uncover the generative process that leads to the observed network. In order to do this, the procedure generates ensembles of random networks in order to estimate how much the observed network departs from the random networks ([Robins, Snijders, Wang, Handcock, & Pattison, 2007](#); [Snijders, Pattison, Robins, & Handcock, 2006](#)). As such, our approach applies similar thinking to the estimation of path-length and clustering coefficient.

The second difference between our framework and the traditional method is that we allow for flexibility in terms of the random network being generated; an analyst can pick the type of random network deemed to be most appropriate to study the network at hand. There are three traditional dimensions used to refer to network types: whether a network is (i) directed or not, (ii) weighted or not, and (iii) whether it is a 1-mode (unipartite) or 2-mode (bipartite) network. Directed networks are networks in which ties have a direction from one node to the next. For example, in the case of a message network, there is a sender and a receiver for each message. Conversely, a collaboration network might be undirected, simply recording that two individuals have worked together. Weighted networks are networks in which a tie is assigned a numerical value to represent a specific attribute. For example, in a collaboration network, the weight might record how often two-individuals have worked together. Finally, 1-mode networks have only 1 type of nodes, while 2-mode networks have two types of nodes. In a 2-mode network, two nodes from the same type can only be linked together through a node from the other type. 2-mode

networks are a convenient way of representing co-presence at events. For example, in a co-authorship network, authors are linked to papers and two authors can only be linked together through having co-authored a paper. We provide an example of our framework for 1-mode unweighted and undirected networks, for 1-mode weighted or directed networks, and for 2-mode weighted networks. To use our framework, an analyst should go through the following five steps:

1. Determine the type of network to be studied (e.g. 2-mode weighted).
2. Based on the type of network studied, choose the most relevant method to calculate the average path-length and the clustering in the network under study (see Table 1 for recommendations).
3. Based on the type of network studied; choose a randomization procedure (see Table 1 for recommendations).
4. Next, generate an ensemble of random networks with the chosen randomization procedure and calculate path-length and clustering in each of those networks. The size of the ensemble should be adapted based on the size of the network under study; the computation is costly and, therefore, there is a trade-off between the time the simulation takes to run and the ability to detect smaller departures from random—obtained through the number of replication. (For convenience, one can use the software we provide with this paper)<sup>ii</sup>.
5. Once the simulation is over, check that the distribution of values for the simulated networks is normal (we recommend visual inspection in a graph and additional diagnostics of normality). If the distribution appears to be normal, the researcher can calculate p-values or plot confidence intervals. For further analyses, to enter the results as variables in a regression for example, we recommend calculating the departure of

the observed path-length and clustering to the mean of the simulation, in number of standard deviations. This has two benefits. First, it contains more information than a dummy coding (of whether the value is higher or lower than random). Second, it facilitates comparisons across different networks.

## Method

Our framework relies on estimating two network characteristics: *average path-length* and *clustering*. There are several ways of estimating a given property depending on the type of network under study; as an example there are different formulas for 1-mode undirected networks and for 2-mode networks for clustering. The randomization procedures can also be adapted, depending on which type of network is being studied. The following section covers the different estimates for the two characteristics of networks we are focusing on—along with the randomization procedures we recommend—based on the types of network under study. Our framework offers the analyst the flexibility to choose the most appropriate type of random network to generate for comparison with the network under study.

**Network Distance.** In binary and weighted 1-mode networks, network distance is the average shortest path-length found among nodes in the largest component. In weighted networks, the shortest path is based on Dijkstra's algorithm ([Dijkstra, 1959](#); [Newman, 2001b](#)). This algorithm identifies the least costly path between two nodes as the shortest path. A higher weight is understood by the algorithm as being a tie that is costlier to traverse. In most networks (and all those analyzed in this paper) the tie weights are positive, and a higher weight represents a stronger tie. Hence we need to transform the weights to represent cost of transmission between two nodes, instead of representing the strength of association. One such transformation process is to invert the

tie weights ([Newman, 2001b](#)). This ensures that a tie that is twice as strong as another (e.g., 4 versus 2) is assigned half the cost (e.g., 0.25 versus 0.5). However, this procedure assigns a non-meaningful distance value to the path. To overcome this issue, we multiply the inverted weights with the average tie weight in the network. Thus, one unit of distance refers to one step with the average tie weight, and distance among nodes becomes comparable across networks. One should note that this transformation does not affect the relative distance within a network, it only acts as a normalization to make comparisons easier (in the rest of this paper, we refer to this procedure as normalizing weights).

In binary 2-mode networks, the distance between nodes is found by projecting the network to a weighted 1-mode network, then applying the method outlined above. To increase the richness of 1-mode projections, it is possible to assign tie weights based on the 2-mode structure ([Newman, 2001b](#); [Opsahl, 2013](#)). We assign a tie weight between two nodes that is equal to the number of common nodes they shared in the 2-mode network. For example, in a scientific collaboration network, the tie strength between two authors would be the number of papers they have co-authored.

To determine distance in weighted 2-mode networks, we extend the projection method by defining the 1-mode tie weight from a node to another as the sum of tie weights towards common nodes in the 2-mode network. As two nodes might exert different levels of interaction with the common nodes, this procedure leads to a weighted 1-mode network where ties are symmetric, but weights are asymmetric. In other words, if A is connected to B, B must be connected to A, but the tie weight from A to B is not necessarily equal to that from B to A. We realize, however, that the most appropriate weight calculation for 2-mode projections are somewhat context dependent. For example, should your 2-mode network record how long individuals have attended a series of

events, you might want the weights to be symmetric and reflect the amount of time the individuals spent together at those event.<sup>iii</sup> Table 1 summarizes the calculation of path-length that we recommend for different networks.

**Network Clustering.** We rely on the global clustering coefficient and its generalizations to weighted 1-mode networks (Opsahl & Panzarasa, 2009) and 2-mode networks (Opsahl, 2013; Piepenbrink & Gaur, 2013) to determine the level of clustering. The global clustering coefficient is similar to the local clustering coefficient, and measures the proportion of triplets that are closed among all the potential triplets in a network. A triplet with nodes A, B, C is closed if ties exist between each pairs (A,B), (B,C) and (A, C). It is formally defined for 1-mode networks as:  $CC = \frac{\text{closed triplets}}{\text{total number of triplets}}$ . It is more robust to skewed degree distribution than the local clustering coefficient. Indeed, the average local clustering coefficient ([Barrat, Barthélemy, Pastor-Satorras, & Vespignani, 2004](#); [Watts & Strogatz, 1998](#)) is disproportionately based on whether ties exist among the neighbors of nodes with few ties ([Opsahl & Panzarasa, 2009](#); [Ravasz, Somera, Mongru, Oltvai, & Barabási, 2002](#)). By contrast the binary global coefficient weights each triplet equally. Another advantage of the global coefficient is that it is faster to compute than the local clustering coefficient. This difference in computation time becomes significant when running simulations with a thousand replications, as we do. Implementations of the clustering coefficient algorithm are readily available in R and Python—in packages such as [igraph \(Csardi & Nepusz, 2006\)](#). In Table 1, we show the method to calculate clustering that we recommend for different types of networks. The first column of Table 1 shows the different types of networks that a researcher might want to analyze; for each type of network, the subsequent columns show the type or types of path-length and clustering measure available. The fourth column shows the randomization procedures available for this type of network and, finally, the fifth column contains recommendations.

---Insert Table 1 here---

**Randomization Procedures.** If the network under study is not a 1-mode unweighted network, using Erdos-Renyi networks for comparison imposes simplification of the features of the observed network. In response, scholars have developed new randomization procedures which maintain additional features of the networks ([Molloy & Reed, 1995](#); [Opsahl \*et al.\*, 2008](#)). We take advantage of those by presenting a framework in which the analyst can select the type of random networks they deem most appropriate for comparison with the network under study.

By reshuffling ties in an observed network, it is possible to create random networks that maintain each node's number of ties (i.e., maintaining the hubs) as well as the overall number of nodes and ties ([Molloy & Reed, 1995](#)). This reshuffling is achieved, for directed networks, by selecting a tie from one node to another and changing the endpoint of that tie. For weighted networks, it is also possible to only reshuffle the weights ([Opsahl \*et al.\*, 2008](#)). A key feature of this procedure is the preservation of the largest interconnected group of nodes. Other methods to refine the generation of random networks have been elaborated, relying on generating functions ([Newman, Strogatz, & Watts, 2001](#); [Newman, Watts, & Strogatz, 2002](#)) and, in some cases, those are less computationally costly than our rewiring procedure. However, they do not always accurately preserve features of a given network, and using them would reduce the generality of our method.

**Weighted 1-mode randomization.** For weighted 1-mode networks, we extend the randomization procedure for random networks to weighted networks by sampling the observed tie weights. Traditionally, classic random networks are binary networks (Erdos & Renyi, 1959). To produce comparison values for weighted networks, a binary random network is far removed from the observed network, and it might be appropriate to use a weighted random network for



comparison. Therefore, we assign tie weights to classic random networks by randomly sampling weights from the tie weights of the network under study. This ensures that the distribution of tie weight in the network—generated in this way—approximate the distribution of weight in the observed network. As opposed to the tie reshuffling ([Molloy & Reed, 1995](#)) and weight reshuffling procedures ([Opsahl \*et al.\*, 2008](#)), this does not maintain the exact weight distribution, as the number of ties in the random network is not exactly the same as the observed network.

**2-mode randomization.** Randomization procedures for 1-mode networks are well established (Erdos & Renyi, 1959; [Molloy & Reed, 1995](#); [Opsahl \*et al.\*, 2008](#)). However, this is not true for 2-mode networks. Our definition of a classic random 2-mode network is a network with a set number of primary and secondary nodes, and a uniform probability of ties being present. When creating a corresponding classic random 2-mode network, the numbers of primary and secondary nodes of the observed network are maintained; the probability of ties being present between nodes from different modes is equal to the number of observed ties divided by the product of the number of primary nodes and the number of secondary nodes ([Barrat \*et al.\*, 2004](#); [Borgatti & Everett, 1997](#)). Similar to our technique for 1-mode random networks, we randomly sample the tie weights in the observed weighted 2-mode network to obtain more realistic random networks. The tie reshuffling procedure for 1-mode networks ([Molloy & Reed, 1995](#)) is extended to 2-mode networks by selecting two nodes from each mode that are connected by exactly two ties and then changing the endpoints. When this procedure has been repeated many times, the resulting network can be considered random, yet the nodes maintain their degree. In our analysis, we repeated the procedure ten times the number of ties in the observed networks. The weight reshuffling procedure for 2-mode networks is a straightforward generalization of the 1-mode procedure ([Opsahl \*et al.\*, 2008](#)). While maintaining the topology of an observed network, it randomly reassigns the tie

weights, ensuring that nodes of both modes maintain their degree, but not their strength.

For all the randomization procedures mentioned there are implementations available in R and Python, through the `igraph` and the `tnet` package ([Csardi & Nepusz, 2006](#); [Opsahl, 2009](#)).

**Determining the number of simulations.** Because each simulation is expensive, one might want to limit the number of runs necessary to accurately estimate the characteristic of a network. We are simulating a distribution of networks and using them to calculate the mean and the standard deviation of this distribution. Intuitively, one understands that the more networks we are simulating the smaller the standard error for the mean of the distribution is going to be. In addition, increasing the runs increases the precision of our estimation of the standard deviation from the mean for this distribution of networks. [Ritter, Schoelles, Quigley, and Klein \(2011\)](#) provide very detailed suggestions on how to calculate the number of runs necessary to detect an effect of a specific size. They suggest that when simulation runs are inexpensive, one could perform 10,000 or 100,000 simulation runs. They also show that, even when differences between the observed mean and the simulated one are relatively small (in the range of 0.2 standard deviation), one requires less than 1,000 runs (the number we have retained here) to obtain a power over 0.99.

### **Description of the networks**

We use six real-world networks to test the proposed method for assessing network efficiency. The networks are (i) the US power grid ([Watts & Strogatz, 1998](#)), (ii) the US airport network ([Colizza, Pastor-Satorras, & Vespignani, 2007](#)), (iii) the neural network of the *C. elegans* worm ([Watts & Strogatz, 1998](#)), (iv) a Facebook-like online social network ([Opsahl & Panzarasa, 2009](#)), (v) a scientific collaboration network ([Newman, 2001a](#)), and (vi) an online forum ([Opsahl, 2013](#)). These six networks represent various combinations of 1- and 2-mode networks, undirected

and directed networks, and binary and weighted networks as outlined in Table 2—which list characteristics of each of those six networks.

### **Power grid**

The first network is the high-voltage power grid in the western states of the United States of America. The nodes are transformers, substations, and generators; the ties are high-voltage transmission lines. This network was also used in the Watts and Strogatz ([Watts & Strogatz, 1998](#)) seminal paper. Although the transmission lines can be directed and differentiated based on their capacity, this information has not been used in previous research and was not available to us. Therefore, we treat this network as an undirected and binary 1-mode network.

### **US Airport**

The US airport network is comprised of the 500 busiest commercial airports in the United States ([Colizza \*et al.\*, 2007](#)). A tie exists between two airports if a flight was scheduled between them in 2002. The tie weights correspond to the number of seats available on the scheduled flights. Although this type of network is directed by nature, as a flight is scheduled from one airport and to another, this network is highly symmetric and we analyze it as undirected. This network is analyzed as an undirected and weighted 1-mode network.

### **C. elegans**

The C. elegans network is the neural network of the *Caenorhabditis elegans* worm, and was included in the Watts and Strogatz analysis (Watts & Strogatz, 1998). The network contains 306 nodes that represent neurons, out of which 297 are part of a single large component (see Table 3). Two neurons are connected if at least one synapse or gap junction exists between them. Given that

not all synapses and gap junctions are bidirectional, this network is directed. Moreover, this network contains information on the number of synapses and gap junctions that transmit signals from one neuron to another. We treat this network as a directed and weighted 1-mode network.

### **Online social network**

The Facebook-like online social network originates from a student online social network website at the University of California, Irvine, which operated in 2004 ([Opsahl & Panzarasa, 2009](#)). This social network was similar to early versions of Facebook in that it only allowed students at a specific university to register and create an online profile. When registered, the students could communicate using private messages or through a forum. This network is constructed from the 59,835 private messages sent on the site. In total, 1,899 users sent or received at least one message. These users are the nodes of the network, and the messages constitute the ties (20,296). As messages are sent from one user to another, this network is treated as a directed and weighted 1-mode network. Although this is the same type of network as the *C. elegans* network, the *C. elegans* network is fairly small (i.e., 306 nodes). We included both networks due to the *C. elegans*' inclusion in the seminal small-world paper (Watts & Strogatz, 1998) and the larger size of the online social network ([Opsahl & Panzarasa, 2009](#)); we aim to demonstrate the applicability of our framework to large networks, yet remain consistent with the existing literature by using a network used in previous studies.

### **Scientific collaboration**

The scientific collaboration network is the co-authorship network based on the 22,016 preprints posted to the Condensed Matter section of the arXiv E-Print Archive between 1995 and 1999 ([Newman, 2001a](#)). This network can be classified as a 2-mode, affiliation, or bi-partite

network since there are two types of nodes (i.e., authors and papers), and connections exist only between the different types of nodes. A tie is formed between an author and a paper if their name appeared on it. This network is a binary 2-mode network as authors cannot be listed more than once on a single paper. Similar to previous work ([Newman, 2001a, 2003](#); [Opsahl, 2013](#)), we deem the authors to be the primary nodes of this network. In so doing, we maintain these nodes when transforming or projecting this network onto a 1-mode network.

### **Online forum**

The online forum network is obtained from the same online social networking website as the online social network ([Opsahl & Panzarasa, 2009](#)); however, the focus in this network is not on the private messages exchanged among users, but on users' activity in the forum ([Opsahl, 2013](#)). The forum represents an interesting 2-mode network among 899 users and 522 topics, as a weight can be assigned to the ties based on the number of messages or characters that a user posted to a topic. In the interest of clarity, we focus on the number of characters. The results are substantively similar when the tie weights are defined as the number of messages. When transforming this weighted 2-mode network into a 1-mode network, we chose to maintain the users, as we believe these are directly responsible for the tie generation. The number of users in this network is smaller than in the online social network; not all users that sent or received private messages participated in the forum.

--- Table 2 about here ---

### **Results**

Table 3 presents comparisons of the six observed networks with the characteristics of 1000 simulated networks. We ran simulations using both transformed networks (following the

traditional method for estimating small-world networks), and non-transformed networks (our approach). Table 3 illustrates the importance of the choice of randomization procedure in obtaining values for comparisons. We ran simulations with transformed versions of the networks using both classic random networks and tie reshuffled random networks; with non-transformed versions of the networks we used classic, tie reshuffled and weight reshuffled random networks. We then estimated the size of the main component for each of the networks (the numbers in brackets present the 95% confidence interval around the average value for the main component obtained from the simulation). Preserving the size of the main component is particularly important as computations of path-lengths are restricted to the main component of a network. As a result, if the main components of the simulated networks are very different in size compared to the one of the observed network, we might obtain estimates for path-length and clustering that do not correspond to the observed network. This is a particularly acute problem for small networks in which the size variations of the main component can have a large incidence on the average path-length. We can see that, in the case of non-transformed networks, the classic random network does not approximate the size of the main component accurately for the US power grid and the scientific collaboration (3<sup>rd</sup> and 4<sup>th</sup> row of table 3). Tie reshuffled random networks performed slightly better, but were still far from the observed values. The same procedures repeated on non-transformed networks show the same limitations for classic random networks. This also illustrates that weight reshuffled networks preserve the main component, making them a good option in cases where other randomization procedures do not reflect the size of the main component accurately. If none of the procedures approximate the main component size accurately, this might indicate that both small-world and efficiency calculations are likely to be unreliable.

---Insert Table 3 about here---

To investigate the efficiency of those six networks, we first assess whether those networks are small-world networks using the existing Watts and Strogatz method (Watts & Strogatz, 1998). As this method necessitates binary and undirected 1-mode networks, we transform the networks and derive approximations of their properties using equations ([Newman, 2001b](#); Watts & Strogatz, 1998). In the top half of Table 4, all network distances are larger (between 8% and 119%), and all clustering levels are higher (between 186% and 105,578%) than approximations based on classic random networks. However, no clear heuristic exists for determining whether these differences constitute comparable network distances and higher clustering-levels; it is a subjective decision of the researchers to deem their network a small-world network ([Langer, Pedroni, & Jäncke, 2013](#)).

---Insert Table 4 about here---

The top half of Table 4 covers the calculation for all the networks, using the small-world framework to estimate properties of networks. All the networks are transformed into 1-mode undirected and unweighted networks to make them comparable with Erdos-Renyi networks. In addition to the path-length and clustering in both classic random networks and tie reshuffled random networks, we also calculate the expected network distance and expected clustering coefficient using equations. We calculate the distance in the observed network, followed by the expected distance using an equation  $\left(\frac{\ln(n)}{\ln(k)}\right)$ , where  $n$  is the number of nodes and  $k$  the average degree (reported as expected network distance and expected clustering coefficient). We then calculate the same parameter in classic random networks (Erdos-Renyi), and finally we calculate this parameter in tie reshuffled networks. We repeat the calculations for clustering coefficient.

We note that the approximations of the network distances using equations are mostly outside of the confidence intervals generated with classic random networks; this highlights the importance of simulation in this specific case. The approximations for clustering coefficient are

much closer to the simulation results of classic random networks. In this first panel, one can see that the network distances in the six networks under study are higher than the ones obtained by using equations. In four cases, the distances in the observed network are higher than the ones obtained with classic and tie reshuffled random networks. However, it is not the case for both the *C.elegans* neural network and the online social network. For the *C.elegans* network, the observed network distance falls within the 95% confidence interval for the simulation using classic random networks, but not when using tie reshuffled networks. With the online social network, the observed distance is outside of the confidence interval for the simulation with classic random networks (but smaller than simulated distances), and within the confidence interval for the simulation using tie reshuffled random networks.

These two results are particularly interesting: they illustrate that the network distance in observed networks is not always similar to the one found in a random network (one of the conditions for a network to be deemed a small-world network). This highlights the importance of defining a heuristic to decide whether or not the observed distance is close to the random distance. Using the Watts and Strogatz framework we show that, of the three networks they used, only the *C.elegans* neural network exhibits distances that are close to distances found in random networks. The results for clustering coefficient indicate that all the networks exhibit clustering levels that are higher than the ones found in random networks, except in the case of the online social network, where the clustering is lower than that found in tie reshuffled random networks.

In summary, all the networks in this test (except the *C.elegans* network), do not seem to exhibit both characteristics of small-world networks at once, when deriving confidence intervals from simulation. Either they do not exhibit path-length that falls within the confidence interval of the simulations using both classic random networks and tie reshuffled random network, or they do



not exhibit clustering that is outside of the confidence intervals for clustering coefficient of both simulations, with classic random networks and tie reshuffled random networks.

Finally, one needs to remark that the practice of computing small-world networks  $Q$  as the ratio of  $\frac{CC_o/CC_r}{PL_o/PL_r}$  would hide that fact; the  $CC_o/CC_r$  ratios in the case of all our networks would always be much larger than  $PL_o/PL_r$ . We believe this is often the case, and limit the usefulness of the small-world  $Q$  as a summary statistic.

Our framework offers the flexibility of using non-transformed networks to estimate both path-length and clustering. In the second panel of Table 4, we use the distribution of values obtained from tie reshuffled and weight reshuffled random networks without transformation. In other words, we use a 2-mode or weighted network when the observed network is 2-mode or weighted.

In this case we obtain different results for network distances, where the simulated network distance confidence intervals show that network distances are similar to the observed network distance in the US airport network, in the online social network, and in the online forum network.

For clustering coefficient, the simulations with non-transformed networks show that clustering in the observed networks is higher than in random networks, in all cases except for the online social network.

Figure 2 illustrates the results from Table 4.

---Figure 2 about here---

Using non-transformed networks, the two networks derived from a Facebook-like online community exhibit network distances comparable to random networks with tie or weight reshuffling ([Opsahl & Panzarasa, 2009](#); [Opsahl, 2013](#)). This may indicate that an online setting

removes barriers for communication with distant parts of the network ([Dodds \*et al.\*, 2003](#); [Kleinberg & Lawrence, 2001](#); [Leskovec & Horvitz, 2008](#)). Moreover, one of these networks also has a level of clustering comparable to that found in random networks. Although online settings might explain the low clustering, it seems that, in this case, this network bears closest similarity to a random network.

Overall, only the US airport network and the online forum network display network distances that are similar to the ones found in non-transformed random networks, and clustering that is higher than that found in random networks. These results show a more nuanced picture of small-world networks than the usual small-world network  $Q$  computation allows for. Indeed, all of the networks tested show considerable differences to random networks, for both clustering and average path-length.

In our framework, a network is deemed efficient if it exhibits clustering coefficient above the one for random networks and outside of the 95% confidence interval of the simulation, while having a path-length falling within the 95% confidence interval of the simulation.

## **Application**

To demonstrate the variations between the methods we propose versus the classic method for measuring small-world properties, we examine the network of collaboration between producers and directors in the French movie industry between 1996 and 2010. This 2-mode network comprises 7567 individuals who participate in 3548 movies. On average, directors and producers participated in 2.24 movies and each movie had 4.77 directors and producers. We chose to focus on this network because collaborative relationships vary substantially during the time period, allowing us to observe changes in path-length and clustering. We use four-year moving windows to build each network (1996-1999, 1997-2000, etc.).

Figure 3 and 4 illustrate the results of calculating clustering and path-length in random networks for each time period, using the classic method and our simulation framework. For the classic path-length, we use binary path-length on the projection of the network, for the classic clustering coefficient, we use the 1-mode clustering coefficient on the projection of the network. The classic method consistently underestimates clustering coefficient compared to our method. Path length estimates using the classic method often fall above the 95<sup>th</sup> percentile of the simulated networks at the beginning of the period, but are within the 5<sup>th</sup> and 95<sup>th</sup> percentile of our simulations for the last three time periods.

--- Insert Figure 3 and 4 about here ---

This example illustrates the difference between the classic method and our estimation framework. Whether the difference in estimation will have a bearing on the comparison between the observed value and the random one depends on the network. In the case of this movie network, the observed distances vary between 6.13 and 4.8, so are continuously higher than the distance in random networks, and the observed clustering varied between 0.044 and 0.077; here again, an order of magnitude above the random clustering results.

Table 5 shows the result of looking at clustering and path-length ratios, using our proposed measure of departure from the mean in number of standard deviations. Looking at the pattern over time, our method allows us to conclude that, both for clustering and path-length, the network behaves less like its random counterparts at the end of the period as it did at the beginning of the period. Using our framework, the analyst can conclude whether the network is converging towards its random counterparts or diverging from them. The path-length and clustering ratios do not offer such a clear picture. In addition, the path-length and clustering

ratios differ by more than an order of magnitude; as such, should the analyst use *small-world*  $Q$  (for example as an explanatory variable in a regression), its value is overly determined by the clustering ratio. In case the analyst wants to use the results of this analysis in a regression, a good starting point would be to include the path-length and clustering departure from random in number of standard deviation as variables in the model.

--- Insert Table 5 about here ---

## **Discussion**

We offer a framework to determine the efficiency of a network, building on the small-world network framework. An efficient network exhibits path-length similar to that of simulated random networks (falling within the 95% confidence interval of the simulation) and clustering above that of simulated random networks (falling above and outside of the 95% confidence interval of the simulation).

Our work alleviates reliance on human judgment, allows probability statements about both clustering and path-length, and offers flexibility in defining values for comparison. Surprisingly, our framework shows that network properties (such as network distances and clustering) do not always behave as expected. More specifically, many networks do not exhibit distances comparable to random networks, and, in some cases, they display clustering values similar to random networks. Table 6 summarizes the relative strength and weaknesses of the two approaches, highlighting that our framework offers flexibility and robustness, at the cost of speed.

--- Insert Table 6 about here ---

Among the networks under study, that have been deemed small-world networks in the past, we cannot conclude that they statistically resemble random networks on path-length and do not

resemble random networks on clustering coefficient. This finding illustrates the importance of rethinking the characterization of small-world networks. Statistical statements regarding the departure from random offer a more precise way of characterizing networks in terms of how efficient they are, with efficiency reflecting their proximity to random on path-length and on clustering coefficient. Those results emphasize the difficulty in answering the question: “Is this network a small-world network?”. They do so by illuminating the heterogeneity among networks, in terms of the proximity of their path-length and clustering to comparable random networks. Far from all having similar relations to their random equivalent, they sometimes show path-lengths that are close to random, and sometimes not. We obtained similar results for clustering coefficient.

Our framework offers several alternative computations to determine whether the path-length or clustering of a studied network are similar to those found in comparable random networks. It allows us to paint a more precise picture of a given network. While a network might be a small-world network when computing small-world network  $Q$ , further investigations using our framework will reveal whether or not network distance and clustering in that network are similar to those found in random networks. It also offers the additional advantage of a probability statement, regarding the likelihood of the observed value to be comparable to one found in a random network (provided the analyst first checks that the distribution of values obtained from random networks is normal). Finally, by defining procedures for classic and non-transformed random networks, we allow scholars flexibility in the features of their networks (such as bi-partite structure and weights) in the calculation, if they deem it relevant.

This study offers a simpler mechanism to estimate the size of the departure from random in a network (e.g., in number of standard deviations). This could lead to the use of more statistically robust measures in modelling the small-world network properties of a network, by replacing small-

world network  $Q$  in regressions with measures of deviation from random. Our results show that small-world network properties are more complex than previously thought. We believe this should ignite further interest in understanding what governs network distance and clustering in real-world networks. Those networks that consistently display characteristics similar to random network distance and higher than random clustering should be of particular interest to researchers. In addition, those networks that are traditionally deemed to be small-world networks, but do not exhibit network distances comparable to random networks, or clustering higher than the one found in random networks, form a new subclass of network. We propose that those networks are less efficient small-world networks, compared to those networks that do pass our test. In so doing, we show that the comparability of real and random network properties is objective, and should not be left to the judgement of individual researchers. Humans show a high propensity to see similarity between two measures, whether or not that similarity is statistically significant (Hand, 2014). Our framework addresses this issue.

### **Recommendations for analysis of path-length and clustering**

Our framework illustrates that the usual practice of calculating small-world network  $Q$  is not always enough to gain an understanding of the path-length and clustering characteristics of a network. We offer a flexible set of tools to help the analyst make probability statements regarding the path-length and the clustering of an observed network. We recommend that, in addition to calculating small-world network  $Q$ , researchers also use our framework to gain a better understanding of how clustering and path-length in their network compare to those found in comparable random networks.

Doing this will provide researchers with additional information about the phenomenon they

are studying that will inform their analyses. If the results of both small-world network calculation and our framework are congruent and indicate the network is a small-world and efficient, the analyst can be confident that the network observed is an efficient small-world network. If the results of the two framework are in opposite direction—with small-world network  $Q$  identifying a small-world network—one is dealing with a non-efficient small-world network. This means that this network is likely to behave differently from the previously identified efficient small-world networks. In the case of the non-efficient small-world network, one expects information to spread slower across the network. If the results of both frameworks are both negative, one can be confident that the network under study will not display accelerated information diffusion in the network. Finally, we have yet to encounter a network that our framework classifies as efficient that does not get classified as a small-world network. We believe such cases to be rare, but such class of networks would be fascinating to study further.

The result of this classification effort in a 2 x 2 matrix will inform further investigation. In the case of either efficient and small-world networks, or non-efficient and non-small-world networks, the researcher can conclude and proceed to further analysis, for example, using our efficiency measure as an independent variable in a regression analysis. Where researchers find it difficult to draw conclusion regarding the properties of a specific network—because it is a non-efficient small-world or an efficient non-small-world—it should prompt them to use related simulation methods, such as exponential random graph models (ERGMs) ([Snijders \*et al.\*, 2006](#)), to study the generative processes which resulted in the observed network. The result from this analysis will help the researcher understand whether to expect the structure of the network to lead to the effects usually associated with small-world networks, such as fast circulation of information. This approach, repeated over a population of networks, would yield interesting insights about

which generative process leads to networks displaying smaller or greater efficiency.

### **Further research and limitations**

We believe our results have considerable consequences for management research. They call for a fresh investigation of networks previously studied by the management literature, such as Canadian banks (Baum, Shipilov, & Rowley, 2003) and strategic alliances ([Schilling & Phelps, 2007](#)) at the firm level, and director interlocks ([Davis \*et al.\*, 2003](#)), inventors ([Fleming \*et al.\*, 2007](#)) and firm co-ownership ([Kogut & Walker, 2001](#)) at the individual level. Indeed, our measures may help explain the difference in the effects of small-world networks reported in those studies. This is an exciting development and should facilitate research into the mechanisms hidden behind the small-world network effect. This exploration, along with research in the contingencies of network effects at different levels (e.g. effects of global structures on the influence of individual position in network on outcomes), are fruitful avenues for future research—our framework will help ensure that results from these future studies rest on solid ground.

Our framework has boundary conditions that are similar to those of the Watts and Strogatz model (1998). It assumes sparse networks in which the number of ties are much smaller than the square of the number of nodes. This limitation should not overly affect the usefulness of our framework, as a large proportion of networks studied in the management literature meet this criterion. One consistent result obtained with the framework is that, in the large majority of cases, clustering is often several orders of magnitude higher in observed networks than in random networks (whether classic or rewired). For this reason, it can lead us to identify networks as small-world networks when computing small-world network  $Q$ , derived from the ratio of the clustering coefficient ratio ( $C_o/C_r$ ) and the path-length ratio ( $PL_o/PL_r$ ). Indeed, this ratio might be large



(the condition for a network to be a small-world network) because the first ratio is always several orders of magnitude bigger than the second, and therefore drives the overall statistic. Retaining only the path-length comparison and using ensembles of random networks to estimate confidence intervals to accept or reject whether a network had a path-length comparable to random networks—seemingly a step in the right direction—would fail to retain important information about the network under study. One area of potentially fruitful exploration would be to compare clustering of the studied networks with clustering in similar regular lattices, instead of random networks. The rationale behind this is that the original assumption in Watts & Strogatz’s model (1998) is not that clustering should be higher than random in small-world networks, but that clustering levels should remain somehow comparable to the ones found in regular lattices with the same number of nodes and same average degree (Watts & Strogatz, 1998).

Another interesting avenue to investigate is the role of hubs in networks ([Schilling & Fang, 2014](#)) and how their presence or absence drives outcomes usually associated with small-world networks. Likelihood and speed of information diffusion in a network might be better approximated by counting hubs and analyzing their connections than by raw path-length ([Schilling & Fang, 2014](#)). Our contribution thus is not only methodological—refining our measurement apparatus—but should also spur new interest in the mapping of theoretical constructs, such as speed of information diffusion, to network measurements.

The classification of networks, and the study of the properties common to a class, is another potentially fruitful avenue for future research. We expect the class of efficient small-world networks to display different properties than non-efficient small-world networks. In turn those networks should display different properties than networks that are neither efficient nor small-world. We hope that future research will explore in-depth the specific properties of those two

classes.

Finally, the definition of a tie in a specific network is going to have an effect on the structure of the network. Sometimes this is obvious, as in the case of ties between airports and between individuals on a social networking website. However, there is potential for further research studying systematic variation across networks representing friendship, advice and other type of relationships.

### **Conclusion**

In this paper, we developed a framework to estimate the efficiency of a network. According to our framework, more efficient networks are likely to sustain faster diffusion of information. Our framework offers insights into why results concerning small-world networks have been mixed and, through classifying networks into two subclasses of efficient and non-efficient small-world networks, offers an avenue for further research to identify the common properties of those two subclasses. The practical advice we offer on how to use the results obtained with our framework as input in further analyses (e.g. regression) should help researchers implement our framework in their own research. We hope this will lead researchers to rethink how they measure small-world network properties in their research.

## REFERENCES

- Albert, R., Jeong, H., & Barabási, A.-L. (1999). The diameter of the world wide web. *Nature*, *401*(September), 130–131.
- Barrat, A., Barthélemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(11), 3747–52.
- Baum, J. A. C., Shipilov, A. V., & Rowley, T. J. (2003). Where do small worlds come from? *Industrial and Corporate Change*, *12*(4), 697–725.
- Borgatti, S. P., & Everett, M. G. (1997). Network analysis of 2-mode data. *Social Networks*, *19*(3), 243–269.
- Burt, R. S. (1992). *Structural hole. The social structure of competition*. Cambridge: Harvard University Press.
- Chung, F., & Lu, L. (2002). The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(25), 15879–82.
- Colizza, V., Pastor-Satorras, R., & Vespignani, A. (2007). Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nature Physics*, *3*(4), 276–282.
- Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Sy*, 1695.
- Davis, G. F., Yoo, M., & Baker, W. E. (2003). The Small World of the American Corporate Elite, 1982–2001. *Strategic Organization*, *1*(3), 301–326.
- Dekker, D., Krackhardt, D., & Snijders, T. A. B. B. (2007). Sensitivity of MRQAP Tests to Collinearity and Autocorrelation Conditions. *Psychometrika*, *72*(4), 563–581.
- Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik*, *1*(1), 269–271.
- Dodds, P. S., Muhamad, R., & Watts, D. J. (2003). An experimental study of search in global social networks. *Science*, *301*(5634), 827–9.
- Erdos, P., & Renyi, A. (1959). On Random Graphs I. *Publicationes Mathematicae Debrecen*, *6*, 290–297.
- Fleming, L., King III, C., & Juda, A. I. (2007). Small Worlds and Regional Innovation. *Organization Science*, *18*(6), 938–954.
- Fleming, L., & Marx, M. (2006). Managing Creativity in Small Worlds. *California Management Review*, *48*(4), 6–27.
- Granovetter, M. S. (2003). Ignorance, knowledge, and outcomes in a small world. *Science*, *301*(5634), 773–4.
- Hand, D. J. (2014). *The Improbability Principle: Why Coincidences, Miracles, and Rare Events*

- Happen Every Day*. Scientific American.
- [Kleinberg, J. M., & Lawrence, S. \(2001\). The structure of the Web. \*Science\*, 294, 1849–50.](#)
- [Kogut, B., & Walker, G. \(2001\). The small world of Germany and the durability of national networks. \*American Sociological Review\*, 66\(3\), 317–335.](#)
- [Langer, N., Pedroni, A., & Jäncke, L. \(2013\). The Problem of Thresholding in Small-World Network Analysis. \*PloS One\*, 8\(1\), e53199.](#)
- [Leskovec, J., & Horvitz, E. \(2008\). Planetary-scale views on a large instant-messaging network. \*Proceeding of the 17th International Conference on World Wide Web - WWW '08\*, 915.](#)
- [Milgram, S. \(1967\). The Small-World Problem. \*Psychology Today\*, 1, 61–67.](#)
- [Molloy, M., & Reed, B. \(1995\). A critical point for random graphs with a given degree sequence. \*Random Structures & Algorithms\*, 6\(2-3\), 161–180.](#)
- [Newman, M. E. J. \(2001a\). Scientific collaboration networks. I. Network construction and fundamental results. \*Physical Review E\*, 64\(1\), 1–8.](#)
- [Newman, M. E. J. \(2001b\). Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. \*Physical Review E\*, 64\(1\), 1–7.](#)
- [Newman, M. E. J. \(2003\). The structure and function of complex networks. \*SIAM Review\*, 45\(2\), 167–256.](#)
- [Newman, M. E. J., Strogatz, S. H., & Watts, D. J. \(2001\). Random graphs with arbitrary degree distributions and their applications. \*Physical Review E\*, 64\(2\), 1–17.](#)
- [Newman, M. E. J., Watts, D. J., & Strogatz, S. H. \(2002\). Random graph models of social networks. \*Proceedings of the National Academy of Sciences of the United States of America\*, 99 Suppl 1, 2566–72.](#)
- [Opsahl, T. \(2009\). \*Structure and Evolution of Weighted Networks\*. University of London \(Queen Mary College\), London, UK.](#)
- [Opsahl, T. \(2013\). Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. \*Social Networks\*, 35\(2\), 159–167.](#)
- [Opsahl, T., Colizza, V., Panzarasa, P., & Ramasco, J. \(2008\). Prominence and Control: The Weighted Rich-Club Effect. \*Physical Review Letters\*, 101\(16\), 1–4.](#)
- [Opsahl, T., & Panzarasa, P. \(2009\). Clustering in weighted networks. \*Social Networks\*, 31\(2\), 155–163.](#)
- [Piepenbrink, A., & Gaur, A. S. \(2013\). Methodological Advances in the Analysis of Bipartite Networks: An Illustration Using Board Interlocks in Indian Firms. \*Organizational Research Methods\*, 16\(3\), 474–496.](#)
- [Rapoport, A., & Horvath, W. J. \(1961\). A Study of a Large Sociogram. \*Behavioral Science\*, 6\(4\), 279–291.](#)
- [Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A.-L. \(2002\). Hierarchical](#)

- organization of modularity in metabolic networks. *Science*, 297(5586), 1551–5.
- Ritter, F. E., Schoelles, M. J., Quigley, K. S., & Klein, L. C. (2011). Determining the Number of Simulation Runs: Treating Simulations as Theories by Not Sampling Their Behavior. In L. Rothrock & S. Narayanan (Eds.), *Human-in-the-Loop Simulations* (pp. 97–116). London: Springer London.
- Robins, G., Snijders, T., Wang, P., Handcock, M., & Pattison, P. (2007). Recent developments in exponential random graph ( $p^*$ ) models for social networks. *Social Networks*, 29, 192–215.
- Schilling, M. A., & Fang, C. (2014). When hubs forget, lie, and play favorites: Interpersonal network structure, information distortion, and organizational learning. *Strategic Management Journal*, 35(7), 974–994.
- Schilling, M. A., & Phelps, C. C. (2007). Interfirm Collaboration Networks: The Impact of Large-Scale Network Structure on Firm Innovation. *Management Science*, 53(7), 1113–1126.
- Simşek, O., & Jensen, D. (2008). Navigating networks by using homophily and degree. *Proceedings of the National Academy of Sciences of the United States of America*, 105(35), 12758–62.
- Singh, P. V. (2010). The small-world effect: The Influence of Macro-Level Properties of Developer Collaboration Networks on Open-Source Project Success. *ACM Transactions on Software Engineering and Methodology*, 20(2), 1–27.
- Snijders, T. A. B., Pattison, P. E., Robins, G. L., & Handcock, M. S. (2006). New Specifications for Exponential Random Graph Models. *Sociological Methodology*, 36, 99–153.
- Steen, J., Macaulay, S., & Kastle, T. (2011). Small worlds: the best network structure for innovation? *Prometheus*, 29(1), 39–50.
- Uzzi, B., Amaral, L. A. N., & Reed-Tsochas, F. (2007). Small-world networks and management science research : a review. *European Management Review*, 4(2), 77–91.
- Uzzi, B., & Spiro, J. (2005). Collaboration and Creativity: The Small World Problem. *American Journal of Sociology*, 111(2), 447–504.
- Verspagen, B., & Duysters, G. (2004). The small worlds of strategic technology alliances. *Technovation*, 24(7), 563–571.
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis, Methods and Applications*. Cambridge: Cambridge University Press.
- Watts, D. J. (2003). *Six Degrees: The Science of a Connected Age*. New York: W. W. Norton.
- Watts, D. J., & Strogatz, S. H. (1998). Collective Dynamics of “Small-World” Networks. *Nature*, 393(4), 440–442.
- Yeates, T. O., & Beeby, M. (2006). Proteins in a small world. *Science*, 314(5807), 1882–3.

## TABLES

**Table 1 Available and recommended measurement for different types of networks**

	Path-length	Clustering	Randomization procedures	Recommended
One-mode network, undirected, unweighted	Unweighted	Global clustering coefficient	Erdos Renyi Tie rewiring	Tie rewiring (preserves degree distribution)
One-mode network, directed, unweighted	Unweighted	Global clustering coefficient	Erdos Renyi Tie rewiring	Tie rewiring (preserves degree distribution)
One-mode network, undirected, weighted	Weighted, normalized weights	Global clustering coefficient	Erdos Renyi Tie rewiring Weight rewiring	Tie rewiring (preserves degree distribution) or weight rewiring (preserves largest connected component, but do not alter structure)
One-mode network, directed, weighted	Weighted, normalized weights	Global clustering coefficient	Erdos Renyi Tie rewiring Weight rewiring	Tie rewiring (preserves degree distribution) or weight rewiring (preserves largest connected component, but do not alter structure)
Two-mode network, undirected, unweighted	Projection, weighted, normalized weights	Two-mode global clustering coefficient	Erdos Renyi Tie rewiring	Tie rewiring (preserves degree distribution)
Two-mode network, directed, unweighted	Projection, weighted, normalized weights	Two-mode global clustering coefficient	Erdos Renyi Tie rewiring	Tie rewiring (preserves degree distribution)
Two-mode network, undirected, weighted	Projection, Asymmetric weights, normalized weights	Two-mode global clustering coefficient	Erdos Renyi Tie rewiring Weight rewiring	Tie rewiring (preserves degree distribution) or weight rewiring (preserves largest connected component, but do not alter structure)
Two-mode network, directed, weighted	Projection, Asymmetric weights, normalized weights	Two-mode global clustering coefficient	Erdos Renyi Tie rewiring Weight rewiring	Tie rewiring (preserves degree distribution) or weight rewiring (preserves largest connected component, but do not alter structure)

**Table 2 Descriptive statistics about the six networks**

	US Power Grid	US Airports	C. Elegans Neural Network	Online Social Network	Scientific Collaborati on	Online Forum
<i>Basic network parameters</i>						
Modes	1	1	1	1	2	2
Directed	No	No	Yes	Yes	No	No
Weighted	No	Yes	Yes	Yes	No	Yes
Primary nodes	4,941	500	306	1,899	16,726	899
Secondary nodes	NA	NA	NA	NA	22,016	522
Ties (one-mode)	6,594	2,980	2,345	20,296	47,594	71,380
Ties (two-mode)	NA	NA	NA	NA	58,595	7,089
Average tie weight	NA	152,320	3.76	2.95	NA	480.19

**Table 3 Comparison of observed networks with random equivalents**

	US Power Grid	US Airports	C. Elegans Neural Network	Online Social Network	Scientific Collaboration	Online Forum
<i>Transformed networks (one-mode, undirected, weights removed)</i>						
Nodes in main component in the observed network	4941	500	297	1,893	13,861	897
% of nodes in main component	100.00%	100.00%	97.06%	99.68%	82.87%	99.78
Nodes in main component in classic random networks	4508.44 [4457; 4557]	499.99 [500; 500]	306.00 [306; 306]	1899 [1899; 1899]	16668.51 [16654; 16682]	899.00 [899; 899]
Nodes in main component in tie reshuffled networks	4752.33 [4711; 4788]	499.65 [498; 500]	296.97 [297; 297]	1894.12 [1897; 1899]	16208.01 [16186; 16228]	897.00 [897; 897]
<i>Non-transformed networks</i>						
Nodes in main component in the observed network	4941	500	239	1,294	13,861	897
% of nodes in main component	100.00%	100.00%	78.10%	68.14%	82.87%	99.78%
Nodes in main component in classic random networks	4,508.44 [4457; 4557]	499.99 [500; 500]	305.75 [304; 306]	1,898.91 [1898; 1899]	16,065.93 [16018; 16115]	898.68 [897; 899]
Nodes in main component in tie reshuffled networks	4,752.33 [4711; 4788]	499.65 [498; 500]	262.36 [257; 266]	1,302.43 [1296; 1308]	15,756.39 [15688; 15821]	897.90 [895; 899]
Nodes in main component in weight reshuffled networks	NA	500.00 [500; 500]	239.00 [239; 239]	1294.00 [1294; 1294]	NA	897.00 [897; 897]

Note: The network distance is the average shortest path-length found among nodes in the largest group of interconnected nodes (i.e., the main or largest component) as distances to nodes in disconnected components are infinite. When randomizing an observed network, the size of the largest components might vary. If the average size of the largest components in random networks is substantially different from the observed network, the comparability of the random networks should be questioned. As this table shows, we find some signal of this. Two issues are worth highlighting. First, the largest components in classic random networks include almost all nodes. This is due to the Poisson distribution of ties across nodes (Erdos & Renyi, 1959), and gives further support for our choice in not using classic random networks to produce comparison values. Second, the size of the largest component does not change when only tie weights are reshuffled (Opsahl et al., 2008). Given that the results shown in Figure 2 and Table 3 are mostly consistent across the two randomization procedures for the weighted networks, we believe the variability in size of the largest component is not a major limitation. Furthermore, the first half of this table repeats this analysis for the transformed networks and their random counterparts. The results of an application of the framework to these networks are reported in Table 4.



**Table 4 Comparison with different type of random networks**

	US Power Grid	US Airports	C. Elegans Neural Network	Online Social Network	Scientific Collaboration	Online Forum
<i>Transformed networks</i>						
Observed Network Distance	18.99	2.99	2.46	3.06	6.63	1.88
Expected Network Distance	8.66	2.51	2.17	2.82	5.59	1.34
Difference	119%	19%	13%	8%	19%	40%
Network Distance in classic random networks	8.48 [8.25; 8.70]	2.77 [2.74; 2.80]	2.46 [2.43; 2.49]	3.09 [3.07; 3.11]	5.79 [5.76; 5.82]	1.82 [1.82; 1.82]
Network Distance in tie reshuffled random networks	8.49 [8.42; 8.57]	2.59 [2.57; 2.61]	2.38 [2.36; 2.40]	3.04 [3.03; 3.06]	4.94 [4.94; 4.95]	1.86 [1.85; 1.86]
Observed Clustering Coefficient	0.10	0.35	0.18	0.06	0.36	0.50
Expected Clustering Coefficient	0.00	0.02	0.05	0.01	0.00	0.18
Difference	18992%	1371%	293%	640%	105578%	186%
Clustering Coefficient in classic random network	0.00 [0.00; 0.00]	0.02 [0.02; 0.03]	0.05 [0.04; 0.05]	0.01 [0.01; 0.01]	0.00 [0.00; 0.00]	0.18 [0.18; 0.18]
Clustering coefficient in tie reshuffled random networks	0.00 [0.00; 0.00]	0.24 [0.23; 0.25]	0.11 [0.11; 0.12]	0.08 [0.08; 0.08]	0.01 [0.01; 0.01]	0.43 [0.43; 0.43]
<i>Non-transformed networks</i>						
Observed network distance	18.99	20.30	8.41	2.96	7.03	1.40
Network distance in classic random networks	8.48 [8.25; 8.70]	2.24 [2.15; 2.34]	3.43 [3.24; 3.64]	3.50 [3.43; 3.57]	4.72 [4.69; 4.76]	0.62 [0.56; 0.69]
Network distance in tie reshuffled random networks	8.49 [8.42; 8.57]	24.32 [5.50; 83.53]	3.46 [3.27; 3.66]	2.96 [2.91; 3.01]	4.10 [4.08; 4.11]	1.04 [0.76; 1.67]
Network distance in weight reshuffled random networks	NA	33.40 [7.05; 109.11]	5.51 [4.88; 6.16]	2.75 [2.70; 2.81]	NA	1.07 [0.84; 1.70]
Observed clustering coefficient	0.10	0.48	0.24	0.06	0.28	0.53
Clustering coefficient in classic random networks	0.00 [0.00; 0.00]	0.02 [0.02; 0.03]	0.03 [0.02; 0.03]	0.01 [0.01; 0.01]	0.00 [0.00; 0.00]	0.11 [0.10; 0.11]
Clustering coefficient in tie reshuffled random networks	0.00 [0.00; 0.00]	0.24 [0.23; 0.25]	0.09 [0.08; 0.10]	0.07 [0.07; 0.07]	0.00 [0.00; 0.00]	0.50 [0.48; 0.51]
Clustering coefficient in weight reshuffled random networks	NA	0.35 [0.34; 0.36]	0.18 [0.17; 0.19]	0.05 [0.05; 0.06]	NA	0.50 [0.48; 0.51]

Note: Empirical results with random networks corresponding to the transformed networks and classic random

networks corresponding to the non-transformed networks. The first part of the table highlight attributes of the analyzed networks. The second part shows results based on Watts and Strogatz’s method (Watts & Strogatz, 1998), it does not take into consideration the recent advances in measuring the network distance and clustering in weighted and/or two-mode networks (M. E. J. Newman, 2001; Opsahl & Panzarasa, 2009; Opsahl, 2013). It should be noted that the approximations are mostly outside the 95% confidence interval of the distributions. This highlights the need for using simulations instead of approximations. The third part of the table shows the result reported graphically in Figure 2 and of classic random networks for the non-transformed networks.

**Table 5 Ratio v. standard deviation departure**

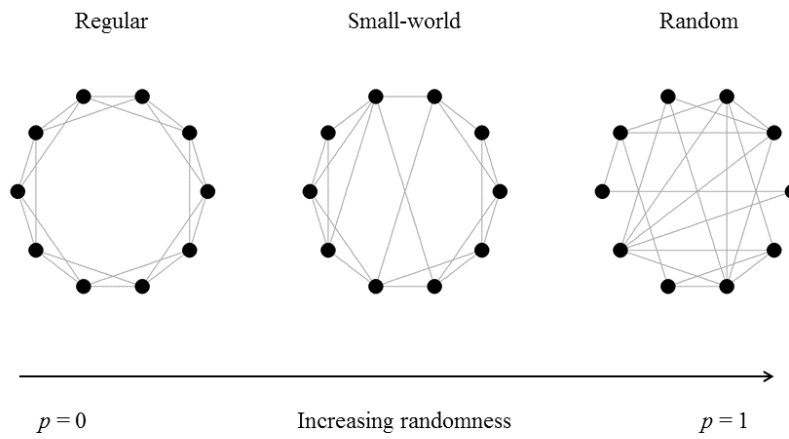
	Clustering ratio	Clustering departure	Path-length ratio	Path-length departure
2000	14.95	47.4	1.33	27.13
2001	17.46	52.61	1.4	38.77
2002	20.37	51.17	1.36	35.87
2003	19.64	43.92	1.29	34.06
2004	21.38	49.86	1.29	37.17
2005	18.53	52.98	1.29	39.56
2006	19.38	68.42	1.24	36.73
2007	23.82	75.71	1.24	32.96
2008	24.81	91.94	1.24	34.97
2009	28.57	101.68	1.26	35.63
2010	21.02	99.86	1.33	48.11

**Table 6 Summary of strength and weaknesses of the two approaches**

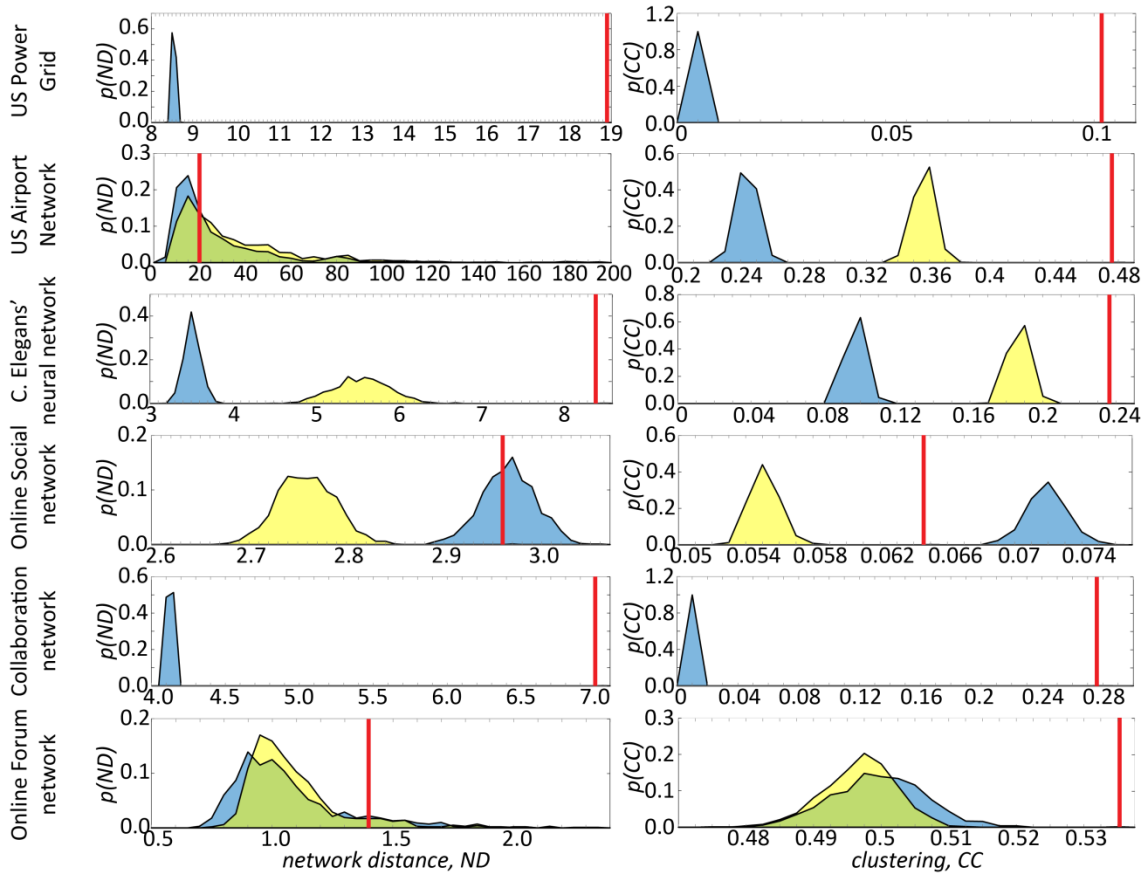
	<b>Traditional Small world estimation</b>	<b>Efficiency estimation</b>
<b>Strength</b>	<ul style="list-style-type: none"><li>• Ease of use</li><li>• Speed</li></ul>	<ul style="list-style-type: none"><li>• Implemented for different types of networks</li><li>• Robust</li></ul>
<b>Weakness</b>	<ul style="list-style-type: none"><li>• Only defined for one-mode undirected and unweighted networks</li></ul>	<ul style="list-style-type: none"><li>• Computationally intensive</li></ul>

## FIGURES

**Figure 1: From ring lattice to random network through rewiring**

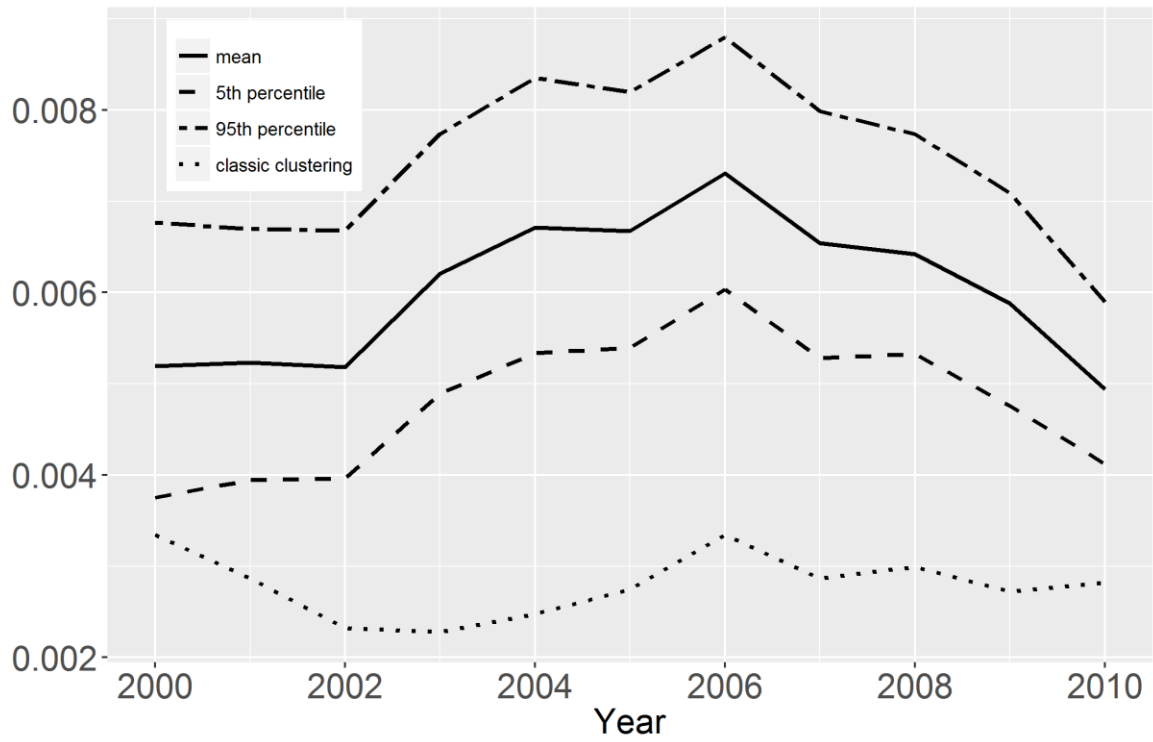


**Figure 2: Empirical analysis of six real-world networks**

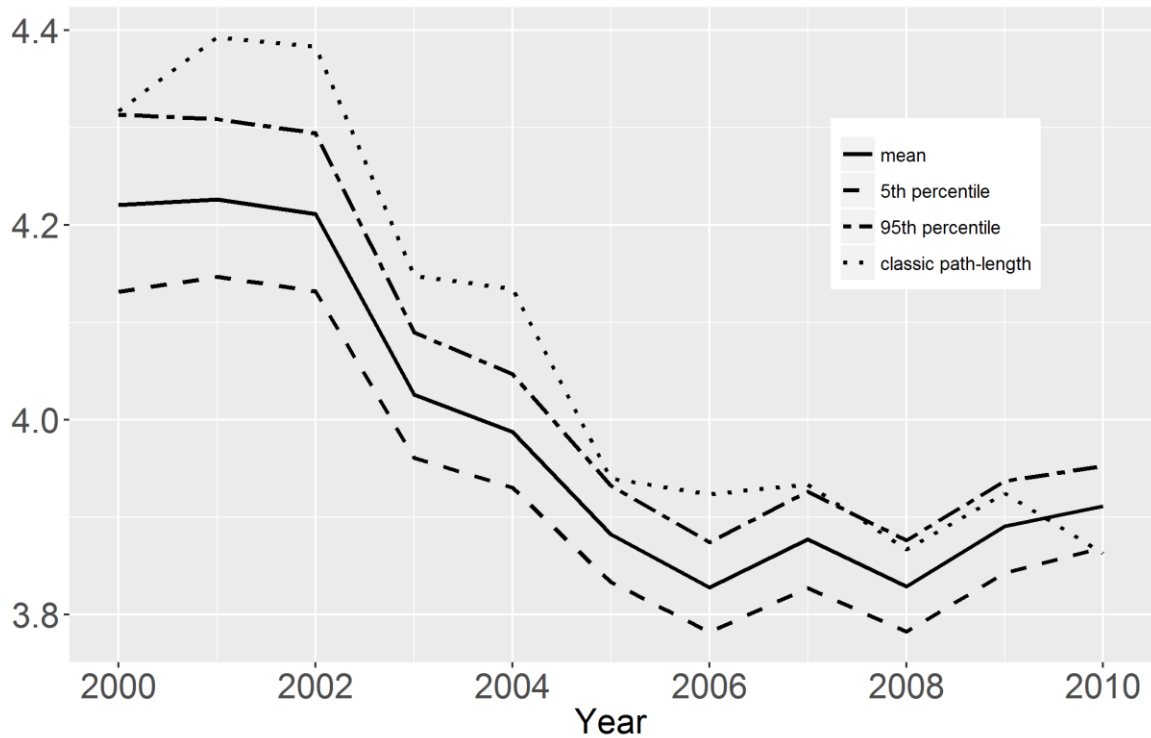


Note: Observed values are shown with red lines, while blue and yellow areas represent distributions of values from tie reshuffled networks and weight reshuffled networks (based on 1000 randomization), respectively. Green areas represent overlap between the distributions. As only weighted networks have corresponding weight reshuffled random networks, we base our conclusions on the results from the tie reshuffled random networks. It is only for the online social network that the results from the randomization procedures do not converge.

**Figure 3: Clustering in random networks obtained from the film network**



**Figure 4: Path-length in random networks obtained from the film network**



#### ENDNOTES

<sup>i</sup> One should note that Burt defines a measure of network efficiency at the individual level, which measures the proportion of one's contacts that are non-redundant (Burt, 1992). Our measure is at the level of the network instead of the individual. However, it is probable that efficient networks following our definition have members with a high average efficiency as defined by Burt.

<sup>ii</sup> The software is available in the R package `tnet`.

<sup>iii</sup> We thank one of the reviewers for suggesting that weighting 2-mode projections is context dependent and one has to allow researchers to pick the approach that is the most appropriate for the analysis they are conducting.