Network Science **5** (1): 30–44, 2017. © Cambridge University Press 2017
doi:10.1017/nws.2017.5

How small is it? Comparing indices of small worldliness

ZACHARY P. NEAL

Psychology Department, Michigan State University, 316 Physics Room 262, East Lansing, MI 4882, USA
(*e-mail:* zpneal@msu.edu)

Abstract

Many studies have attempted to determine whether an observed network exhibits a so-called "small-world structure." Such determinations have often relied on a conceptual definition of small worldliness proposed by Watts and Strogatz in their seminal 1998 paper, but recently several quantitative indices of network small worldliness have emerged. This paper reviews and compares three such indices—the small-world quotient (Q), a small-world metric (ω), and the small-world index(SWI)—in the canonical Watts–Strogatz re-wiring model and in four real-world networks. These analyses suggest that researchers should avoid Q, and identify considerations that should guide the choice between ω and SWI.

Keywords: *small world, clustering, path length*

1 Introduction

The notion that the social world is "small" is an old one: Information can spread rapidly because individuals are separated by a surprisingly small number of intermediate links (Milgram, 1967). The smallness of the world has seemed paradoxical, however, because from the perspective of each individual the world appears relatively local and clustered, for example, in social circles organized by geography or interest. The phenomenon of small worlds was poorly understood until Watts & Strogatz (1998) (hereafter WS) identified the network structure that makes it possible. Since their discovery of an empirically common class of networks—small-world networks—there has been significant interest in identifying real-world networks that exhibit this structure. Multiple indices have been proposed that aim to capture whether or not, or to what extent, a given network has such a structure, but it remains unclear which indices are helpful and meaningful in making such decisions. The goal of this paper is to compare and evaluate these indices of network small worldliness.

The most widely used definition of small-world structure is a conceptual or qualitative one: A network has a small-world structure when its clustering coefficient (C, see materials and methods) is relatively high, but its mean path length (L) is relatively low (WS). This conceptual definition derives from the fact that a high clustering coefficient accounts for the phenomenon of social groups, while the short mean path length accounts for the phenomenon of rapid information spread and of "six degrees of separation." The challenge in applying this definition lies in specifying: relative to what, and how high (low) is high (low) enough? In the initial

https://doi.org/10.1017/nws.2017.5 Published online by Cambridge University Press

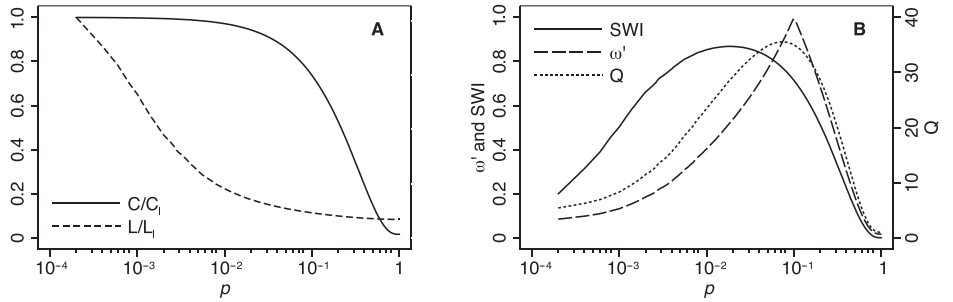


Fig. 1. Comparison of small-worldliness indices in a Watts–Strogatz model. All edges in a lattice with $n = 1000$ and $k = 10$ were randomly rewired with probability p (note the log-scale on the x-axis); all values represent averages over 25 realizations. (A) In the WS model, when $0 < p \ll 1$, the network has a mean path length that is similar to a corresponding random graph, but a clustering coefficient that is similar to a corresponding lattice. (B) The three indices seek to quantify the extent to which the network is small worldly; their values are shown for each value of p. Note that ω' and SWI range between 0 and 1 (left axis), while Q's range depends on n and k (right axis).

WS formulation, this meant having a clustering coefficient similar to that of a one-dimensional ring lattice, and a mean path length similar to that of an Erdös–Rényi random graph, but alternate reference graphs can also be used to decide when C is sufficiently high and L is sufficiently low. The WS model demonstrated that such a structure emerges when a small number of edges in a lattice are randomly rewired (see Figure 1(A)).

1.1 Why measure small worldliness?

Before turning to indices of small worldliness, it is worth asking why it is useful to measure a network's small worldliness. Since WS introduced the notion of small worldliness to network science, it has become almost as common for empirical studies of networks to report whether the network under consideration has a small-world structure as to report other basic descriptives like size and density. This trend toward viewing a network's small worldliness as a basic descriptive statistic can be seen across multiple contexts including transportation networks (Guimera et al., 2005; Neal, 2014; Bagler, 2008; Han et al., 2009; Cardillo et al., 2006; Sen et al., 2003), neural networks (Telesford et al., 2011; Humphries et al., 2006; Bassett et al., 2008; Guye et al., 2010; Liu et al., 2008; Bullmore & Sporns, 2009), biological networks (Wagner & Fell, 2001), corporate networks (Kogut & Walker, 2001; Davis et al., 2003), and human and animal social networks (Uzzi & Spiro, 2005; Lusseau, 2003; Guimera et al., 2003). Indeed, since 2010, more than 600 papers have appeared that contain the phrase "...is a small world network." However, in many of these cases, the claim of small worldliness relies only the conceptual definition, which lacks precision. This approach may be suitable in cases where the network's small worldliness is clear (i.e. when C is very high and L is very small), but is less helpful when a network's small worldliness is ambiguous. Because small worldliness has come to be viewed as a basic descriptive property of a network, a standard approach to measure small worldliness that could accompany standardized ways to

https://doi.org/10.1017/nws.2017.5 Published online by Cambridge University Press

report other descriptives like density would bring greater precision, consistency, and comparability across studies.

Even if small worldliness *can* be measured, this does not necessarily imply that it *should* be measured. The past several years have seen a proliferation of network metrics, and some have argued that focusing only on topological characteristics can obscure important substantive differences between networks (Grannis, 2010; Neal, 2014). However, an emerging body of empirical and theoretical work suggests small worldliness has significant conceptual value, and that an index (or indices) of small worldliness would be useful for testing theoretical propositions about its effects. For example, empirical studies have found that a network's small worldliness is associated with artists' creativity in Broadway musical productions (Uzzi & Spiro, 2005), with the technical and commercial success of open-source software (Singh, 2010), and with firms' patent output across 11 different industries (Schilling & Phelps, 2007). Others have hypothesized that "there are entrepreneurial strategies through which new ventures might achieve the advantages of small-world networks (Aldrich & Kim, 2007), that small-world networks can help explain information transfer in the brain (Gallos et al., 2012), and that small worldliness can provide a solution to the research-practice gap in social services (Neal et al., 2015). Finally, still others have identified small worldliness as a way to capture social phenomena that are difficult to measure, including collaboration potential in community change efforts (Lawlor & Neal, 2016) and social capital in urban neighborhoods (Neal, 2015). These past empirical studies illustrate that indices of small worldliness have already been useful, while small-world-based hypotheses and measurement strategies proposed in more theoretical work highlight the need to understand these indices better.

1.2 Indices of small worldliness

Indices of small worldliness are computed from the observed network's clustering coefficient (C) and mean path length (L). The principal difference among indices is how these values are normalized with respect to their values in one or more corresponding reference graphs. In the presentation of each index below, C_l and L_l refer to the clustering coefficient and mean path length in a lattice reference graph characterized by a high C and L. Similarly, C_r and L_r refer to the clustering coefficient and mean path length in a random reference graph characterized by a low C and L. After presenting the three indices, I briefly discuss the selection of specific lattice and random reference graphs.

Single-graph normalized index, Q. The earliest index of a network's small worldliness normalized both C and L with respect to a single reference graph:

$$Q = \frac{C/C_r}{L/L_r} \tag{1}$$

It was initially presented without a name (Kogut & Walker, 2001), but later was called the small-world quotient (Q) (Davis et al., 2003; Uzzi & Spiro, 2005) and a measure of "small-world-ness" (S) (Humphries et al., 2006). This index has become widely adopted in the literature (Bassett et al., 2008; Guye et al., 2010; Liu et al., 2008; Bullmore & Sporns, 2009). Values of $Q > 1$, which can be identified using an

https://doi.org/10.1017/nws.2017.5 Published online by Cambridge University Press

associated statistical test (Humphries & Gurney, 2008), are interpreted as evidence of small worldliness because they indicate that the network's mean path length is similar to that of a random reference graph and/or that the network's clustering coefficient is larger than that of a random reference graph. The maximum value of Q depends on the network's size and mean degree. Although it has not been used previously, a related measure that is proportional to Q can also be defined by normalizing C and L with respect to a lattice reference graph:

$$Q' = \frac{L_l/L}{C_l/C} \tag{2}$$

These indices can be classified as single-graph normalized indices because they normalize the network's clustering coefficient and mean path length with respect to only a single reference graph: either a random graph for Q or a lattice for Q'. Although this approach is appealing in its simplicity, it has a couple of drawbacks (Telesford et al., 2011). First, their maximum value depends on the network's size and mean degree, limiting their utility in comparing the small worldliness of networks of differing sizes and degree distributions. This property is illustrated by application to a network of biomedical research co-authorship in Table 2 below. Second, comparing both the network's clustering coefficient and mean path length to the same reference graph is inappropriate because these values in a small-world network are supposed to mimic those in *different* reference networks. In a network with small-world structure, C is not merely expected to deviate from C_r, but rather to approach C_l. By comparing C to C_r, Q captures how different an observed network is from a random reference graph, but not how similar it is to a lattice reference graph, and thus fails to capture the property of interest for detecting small worldliness. Likewise, in a network with small-world structure, L is not merely expected to deviate from L_l, but rather to approach L_r. By comparing L to L_l, Q' captures how different an observed network is from a lattice reference graph, but not how similar it is to a random reference graph, and thus also fails to capture the property of interest for detecting small worldliness.

Double-graph normalized index, ω. Seeking to overcome many of the limitations of a single-graph normalized index, more recent indices have involved normalizing values using *both* random and lattice reference graphs. The first such double-graph normalized index, ω (Telesford et al., 2011), is defined as

$$\omega = \frac{L_r}{L} - \frac{C}{C_l} \tag{3}$$

which ranges between -1 when the network lattice characteristics to $+1$ when the network has random graph characteristics, with values near 0 interpreted as evidence of small worldliness. In this paper, to facilitate comparison with other indices, I use a transformed version:

$$\omega' = 1 - |\omega| \tag{4}$$

which ranges between 0 and 1. Larger values of ω' indicate greater small worldliness, while both random and lattice networks have values near 0.

Compared to Q, ω offers a number of advantages. First, it has a fixed scale that is independent of network size and mean degree, which facilitates comparison across

https://doi.org/10.1017/nws.2017.5 Published online by Cambridge University Press

different networks. Second, it more appropriately compares the observed network's clustering coefficient to that in a lattice reference graph, but the observed network's mean path length to that in a random reference graph. However, in cases where the difference between the clustering coefficient or mean path length in random and lattice reference graphs is small, ω and ω' may nonetheless be misleading. Consider a network where $C = 0.4$ and $L = 3$, and where the random reference graph has $C_r = 0.399$ and $L_r = 2.90$ and the lattice reference graph has $C_l = 0.41$ and $L_l = 3.01$. In this case, $\omega' = 0.99$, which is very near one and thus suggests that the network is characterized by small worldliness. It achieves such a high value because C is very near C_l and L is very near L_r. However, closer inspection reveals that while this is true, C is actually closer to C_r and L is actually closer to L_l, which is the opposite of the conceptual definition of a small-world network. This occurs because, although ω and ω' normalize C and L using values from the correct reference graphs, these normalizations do not take into account the amount of possible variation between C_r and C_l, or between L_r and L_l. This property is illustrated by application to an E.Coli metabolic network in Table 2 below.

Double-graph normalized index, SWI. A more recent double-graph normalized index, the small-world index (SWI), is defined as

$$SWI = \frac{L - L_l}{L_r - L_l} \times \frac{C - C_r}{C_l - C_r} \tag{5}$$

and aims to overcome the limitation of ω by capturing the location of the observed network's clustering coefficient and mean path length between the extreme values observed in corresponding random and lattice reference graphs (Neal, 2015). The term on the left ranges from 0, when $L = L_l$ which is not characteristic of a small-world network, to 1, when $L = L_r$ which is characteristic of a small-world network. Similarly, the term on the right ranges from 0, when $C = C_r$ which is not characteristic of a small-world network, to 1, when $C = C_l$ which is characteristic of a small-world network. Thus, the product of these two terms ranges from 0, when the network displays only one or neither small-world characteristic, to 1, when the network displays both small-world characteristics.[1] It therefore aims to measure the extent to which an observed network's two small-world characteristics—clustering coefficient and mean path length—are jointly maximized. This index most closely matches the WS conception of small worldliness as occurring along a continuum between lattice and random reference graph, with the network's clustering coefficient nearer the lattice extreme but its mean path length nearer the random extreme.

1.3 Picking reference graphs

These indices define classes of small-worldliness metrics, but computing them requires specifying the properties of the reference graph(s) to be used. The lattice reference graph must be characterized by a relatively high C and L, the random reference graph must be characterized by a relatively low C and L, and both must have the

[1] Sporns & Zwi (2004) proposed the same approach for normalizing C, but the inverse for L (i.e. $\frac{L-L_r}{L_l-L_r}$) and did not combine the standardized metrics into a single index.

https://doi.org/10.1017/nws.2017.5 Published online by Cambridge University Press

same size and density as the observed network. Beyond these basic requirements, the properties of the reference graphs may be further constrained to match characteristics of the observed network. For example, the researcher may wish to require that the reference graphs preserve the observed network's degree distribution. Such considerations are particularly important when the observed network comes from a context that imposes constraints on its structure. For example, many attempts to identify real-world networks with small-world characteristics have looked to transportation infrastructure networks including roads and rail, which are generally constrained as planar graphs (Sen et al., 2003; Cardillo et al., 2006). When computing indices of small worldliness for these networks, it would be most appropriate to constrain the reference graphs to also be planar (Masucci et al., 2009). Similarly, if the observed network contains structural ones or zeros (i.e. certain edges that are required or prohibited), these features should be preserved in the reference graphs.

For the purposes of computing and comparing these indices below, the lattice reference graph is a one-dimensional ring lattice and the random reference graph is an Erdös–Rényi random graph. Both reference graphs have the same size, density, and mean degree as the observed network. Although other reference graphs are possible, these are used for three reasons. First, these are the reference graphs used in the original WS model. Second, all three indices under consideration were originally developed with these reference graphs in mind, and all subsequent applications of the indices have used these reference graphs. Finally, the computation of their clustering coefficients and mean path lengths is simplified because these values are well approximated in sparse graphs as

$$C_r = \frac{k}{n} \tag{6}$$

$$L_r = \frac{ln\ n}{ln\ k} \tag{7}$$

$$L_l = \frac{n}{2k} \tag{8}$$

$$C_l = \frac{3(k-2)}{4(k-1)} \tag{9}$$

2 Results

2.1 Comparison in the Watts–Strogatz model

The WS model of small-world network formation begins with a lattice of n nodes, in which each is connected to its k nearest neighbors (see materials and methods). All edges are then randomly rewired with probability p. When $p = 0$, the network remains a lattice, while when $p = 1$, the resulting network is a random graph. As p increases from 0, the network's mean path length (L) rapidly declines and approaches its final value in a random network, but the network's clustering coefficient (C) initially remains stable (WS) (see Figure 1(A)). The networks that occur for intermediate values of p, where the solid and dashed lines are most divergent—that is, which retain a high lattice-like clustering coefficient, but have a small random-like mean path length—are small-world networks.

https://doi.org/10.1017/nws.2017.5 Published online by Cambridge University Press

Table 1. *Maximum small-worldliness index values in the Watts–Strogatz model.*

Index	Maximum	p	Edges rewired
Q	35.547	0.073	366
ω'	1.000	0.097	487
SWI	0.868	0.018	88

This model offers a standardized framework for comparing indices of small worldliness, and for ensuring that each index behaves as required by a conceptual definition of small worldliness (Humphries & Gurney, 2008; Telesford et al., 2011). Figure 1(B) shows the value of SWI (solid line), ω' (dashed line), and Q (dotted line) for each value of p. Each index behaves as required: A skewed inverse-U relationship with p such that the index has a low value when $p = 0$ or 1, and reaches its maximum value at a relatively small p. However, despite exhibiting behavior that mirrors expectations from the WS model, Table 1 shows that these indices differ in their maximum values and the p at which their maximum value is achieved.

Maximum value. In a network with $n = 1000$ and $k = 10$, Q achieves a maximum value of 35.547. However, any given network's Q is a function of n and k. A comparison of 30 real-world networks deemed to have small-world structure because $Q > 1$ had values ranging from 1.67 (where $n = 459$ and $k = 6.2$) to 61967 (where $n = 1520251$ and $k = 15.53$) (Humphries & Gurney, 2008). In principle, a network's Q could be normalized as a function of the maximum possible value for a network with the same n and k; however, this is impractical because it requires a computationally intensive search for the maximum Q. In the absence of a practical approach to normalization, Q is of limited use for comparing the small worldliness of different networks.

Unlike Q, ω' and SWI both range between 0 and 1, which makes them useful for comparing the small worldliness of networks with differing characteristics including different n and k. However, in the WS model, ω' always achieves its maximum value, while SWI never achieves its maximum value. To understand why this happens requires considering differences in the meaning of a value of 1 for these two indices. For ω', a value of 1 indicates that "the network is balanced such that path length is as close to random as clustering is to a lattice" (Telesford et al., 2011, p.370). As Figure 1(B) illustrates, in the transition from a ring lattice to a random graph in a traditional WS model, such a network always exists. An observed network with $\omega' = 1$ is as small worldly as *mathematically* possible, given any constraints imposed on the reference graphs (here, only n and k). Smaller values of ω' indicate how close an observed network is to this mathematical ideal. In contrast, for SWI, a value of 1 indicates that the network's clustering coefficient matches that in a lattice reference graph and its mean path length matches that in a random reference graph. This describes a network that is as small worldly as *conceptually* possible, that is, that perfectly matches the WS conceptual definition. However, as Figure 1(B) illustrates, $SWI \neq 1$ for any p because such a network does not exist. Thus, values of SWI indicate

https://doi.org/10.1017/nws.2017.5 Published online by Cambridge University Press

how close an observed network is to this conceptual, but unachievable, ideal. This property is illustrated by application to the US power grid network in Table 2 below.

Randomness to achieve maximum value. In the WS model, p indexes the amount of randomness in an otherwise highly ordered and clustered network. The highly clustered "large world" at $p = 0$ transitions into a highly connected, yet still clustered "small world" when $0 < p \ll 1$. That is, introducing only a small amount of randomness to a lattice is necessary to generate a small-world network. This occurs because the first few random edge rewirings have a highly non-linear effect on L, while they have very little effect on C. A network in this model should achieve its maximal small worldliness at the p where L's rapid decline tapers off but before C begins to decline appreciably. The three indices under consideration achieve their maximum values at different p, and thus differ in the amount of randomness they view as necessary to maximize a network's small worldliness.

The SWI views the most small worldly $n = 1000/k = 10$ network as occurring after about 2% of edges have been randomly rewired from an initial lattice configuration. At this p, L has declined 85.3% from 47.7 to 7.029, but C has declined only 4.8% from 0.695 to 0.661. This is the greatest difference between the percent decline in L and C (i.e. the distance between the curves in Figure 1(A)) for all p.

In contrast, Q and ω' both view the most small-worldly network as occurring much later, after around 7% to 10% of edges have been randomly rewired. In this range of p, L has declined 90.5% from 47.7 to about 4.5, and C has declined 24.4% from 0.695 to about 0.525. This represents a slightly shorter mean path length than when $p = 0.02$, but a dramatically smaller clustering coefficient. The networks obtained when $0.07 < p < 0.10$ are certainly still small worldly, but they are not as small worldly as the network obtained when $p \approx 0.02$.

2.2 *Indices' relationships with n and k*

Comparison in the WS model suggests that Q, ω', and SWI may all be viable indices of network small worldliness; however, this model focuses on a single network of $n = 1,000$ and $k = 10$. In this section, following these previous investigations of other small-world indices (Humphries & Gurney, 2008; Telesford et al., 2011), I explore how these indices behave in networks of varying sizes and degrees.

As expected, Q's behavior depends heavily on both n and k (see Figure 2(A) and (D)). Specifically, Q achieves systematically larger values for all p in larger networks, and in sparser networks. Moreover, although Q's lower bound is 0, it has no upper bound. These features highlight why Q cannot readily be compared across networks that vary on these characteristics: Regardless of their actual small worldliness, larger and sparser networks will have larger values of Q than smaller and denser networks.

The behavior of ω' (see Figure 3(B) and (E)) is remarkably stable over all p with respect to n and k, which mirrors earlier findings (Telesford et al., 2011). This stability is expected because ω' always achieves a value of 1 for some p in a WS model.

Finally, the behavior of SWI (see Figure 3(C) and (F)) is stable for networks of varying degree, as well as for networks of $n \geqslant 500$. However, in larger networks,

https://doi.org/10.1017/nws.2017.5 Published online by Cambridge University Press

https://doi.org/10.1017/nws.2017.5 Published online by Cambridge University Press

Table 2. *Small worldliness of real-world networks.*

Network*	n†	k	C_l	C	C_r	L_l	L	L_r	Q	ω'	SWI
Metabolic	315	56.60	0.737	0.220	0.180	2.7826	2.620	1.425	0.6661	0.755	0.009
Airports	407	10.39	0.670	0.527	0.026	19.586	2.960	2.567	17.903	0.919	0.760
Power	4941	2.67	0.301	0.080	>0.001	925.28	18.7	8.661	68.564	0.803	0.262
Biology	1520251	15.53	0.698	0.600	>0.001	48945	4.94	5.190	61955	0.804	0.859

*E.Coli. Metabolic (Humphries & Gurney, 2008; Wagner & Fell, 2001), US Airports (Neal, 2014), US Power Grid (Watts & Strogatz, 1998), Biology Co-authorship (Newman, 2001).

†Values for n (network size), k (mean degree), C (clustering coefficient), and L (mean path length) were obtained from previous work as referenced. Values for clustering coefficients and mean path lengths in corresponding lattice and random graphs (C_l, C_r, L_l, and L_r) and values for small-world indices (Q, ω', and SWI) were computed using the equations presented above.

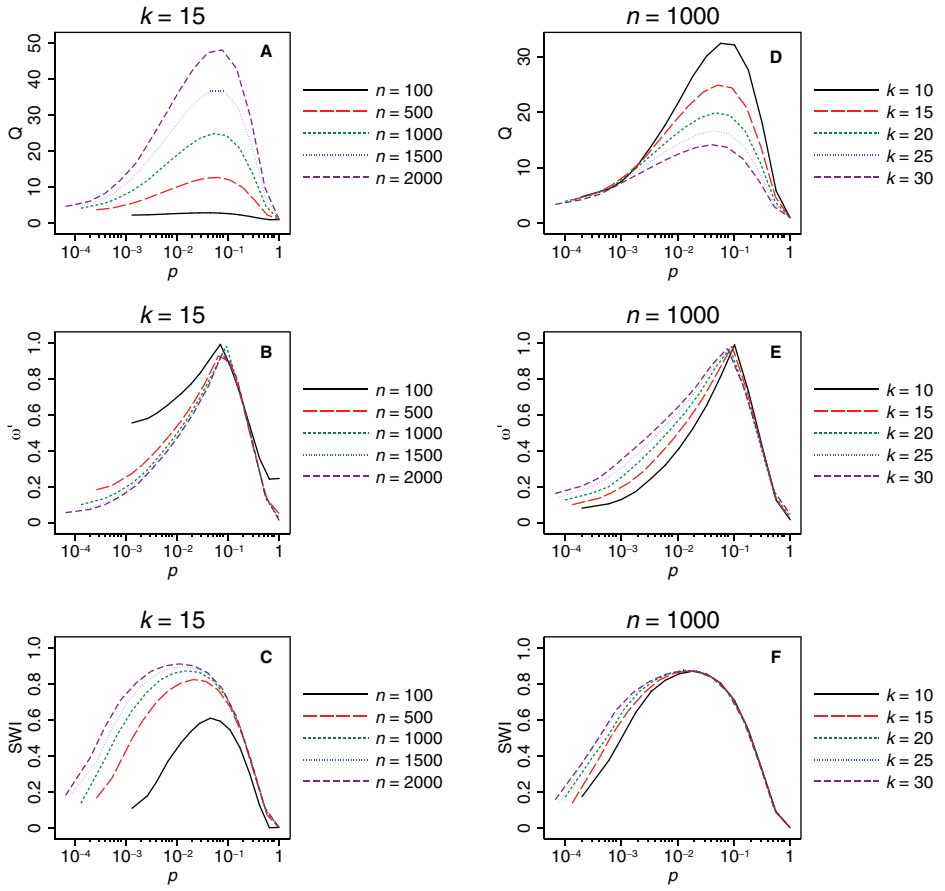


Fig. 2. Comparison of small-worldliness indices over p for networks of varying size (n) and mean degree (k). (A–C) The behavior of ω is stable for networks of differing sizes, but larger networks exhibit systematically higher values of Q and SWI over all values of p. (D–F) The behavior of ω and SWI is stable for networks of differing mean degree, but sparser networks exhibit systematically higher values of Q over all values of p. All values represent averages over 25 realizations. (Color online)

it achieves a larger maximum value, which occurs at lower p. To more closely investigate this phenomenon, Figure 3 illustrates how n, k, and k/n (link density) influence the magnitude and location (with respect to p) of SWI's maxima in a WS model. First, I find that the maximum value of SWI is strongly positively related with n ($r = 0.833$), weakly positively related with k ($r = 0.07$), and strongly negatively related with link density ($r = -0.899$). That is, large and sparse networks have a greater capacity to approach maximal small worldliness (from the perspective of SWI) than small and dense networks. It is important to note that this finding is different from the fact, noted above, that Q is always larger in larger networks. Values of Q grow larger as n increases simply because larger networks tend to have longer mean path lengths. In contrast, the relationship between $max(SWI)$ and n seen in Figure 3(A) highlights a property of small-world structure, namely, that its conceptual ideal requires a relatively large network to appear. Second, I find that the p at which SWI achieves its maximum value is strongly negatively related with n

https://doi.org/10.1017/nws.2017.5 Published online by Cambridge University Press

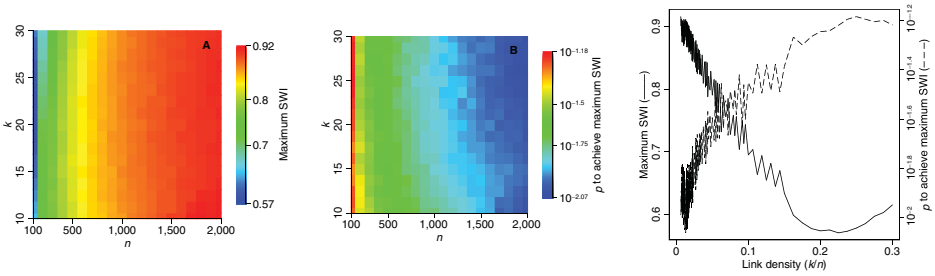


Fig. 3. Magnitude and location of SWI's maxima in Watts–Strogatz models. (A) The maximum value that SWI achieves in a WS model increases with n, but is less influenced by k. (B) The value of p when SWI achieves its maximum value in a WS model decreases with n, and decreases slightly with k. (C) In sparser networks, SWI achieves a larger maximum value, and achieves is maximum value with fewer random rewirings. All values represent averages over 25 realizations. (Color online)

($r = -0.927$), weakly negatively related with k ($r = -0.245$), and strongly positively related with link density ($r = 0.773$). That is, large and sparse lattices transition into small worlds earlier (i.e. with fewer random re-wirings) than small and dense lattices.

2.3 Comparison in real-world networks

In this section, I compare the three small-world indices—Q, ω', and SWI—in four real-world networks that range in size and come from multiple contexts that have previously been examined in the small-world literature. Table 2 reports each network's size (n) and mean degree (k), the values of the three SWIs, and each component of these indices. These comparisons serve as a concrete illustration of the indices, and each network highlights a different property of the indices identified by the more abstract analysis above.

Air traffic networks are frequently cited as contexts where small-world structures are observed (Guimera et al., 2005; Bagler, 2008; Han et al., 2009). The US airport network described in Table 2 closely matches a conceptual definition of small worldliness, exhibiting a clustering coefficient (0.527) that approaches the value observed in a lattice (0.670) and a mean path length (2.96) that approaches the value observed in a random graph (2.567). The values of all three indices – $Q > 1$, ω', and $SWI \gg 0$—suggest this network is small worldly. This example illustrates that in cases where a network's small worldliness is clear given the conceptual definition, all three indices agree and yield expected results.

The network of co-authorship in biomedical research in the MEDLINE database is larger than the airport network by several orders of magnitude, but also exhibits a very clear small-world structure: Its clustering coefficient (0.6) is similar to that of a lattice (0.698), while its mean path length (4.94) is similar to that of a random graph (5.19). Again, the values of all three indices accurately identify this network's small worldliness. However, this network offers an illustration of the effect of n and k on Q, and thus why Q cannot be compared across networks of different sizes or degrees. The co-authorship network's Q is more than 3,000 times larger than in the airport network, but this does not mean the co-authorship network is more than

https://doi.org/10.1017/nws.2017.5 Published online by Cambridge University Press

3,000 times more small worldly than the airport network. Instead, this large value occurs simply because $C_r \approx 0$ in sparse networks like this one, and thus C/C_r (i.e. the numerator in Equation (1)) yields very large values (Telesford et al., 2011).

The E.Coli metabolic network offers a different scenario because it is *not* small worldly. Its mean path length (2.620) is nearly as large as that of a lattice (2.783), while its clustering coefficient (0.220) is nearly as small as that of a random graph (0.180). These values suggest that this network is more than simply non-small worldly, but indeed is the very opposite of small worldly. The $Q < 1$ and $SWI \approx 0$ are consistent with this; however, the $\omega' \gg 0$ incorrectly identifies the network as small worldly. This example illustrates a case where ω' is misleading because it fails to consider both lattice and random reference graphs when normalizing C and L. Because ω' normalizes C only with respect to C_l, it misses the fact that in this network $C \approx C_r$. Likewise, because ω' normalizes L only with respect to L_r, it misses the fact that in this network $L \approx L_l$.

Finally, the US power grid network has been regarded as an example of small-world structure since its initial use as an example by WS. Consistent with the conceptual definition of small worldliness, this network's clustering coefficient (0.08) is much larger than that of a random graph ($C_r > 0.0001$), and its mean path length (18.7) is similar to that of a random graph ($L_r = 8.66$). Both Q and ω' yield values that confirm its small worldliness, but an SWI of only 0.262 suggests it is not particularly small worldly. This illustrates a case where SWI's comparison of this network to a conceptual ideal is too stringent. The small SWI value occurs because, although the network's clustering coefficient is larger than that of a random graph, it is still not close to that of a lattice. Specifically, $\frac{C-C_r}{C_l-C_r}$ (i.e. right term in Equation (5)) equals 0.26, which indicates that the observed network's clustering coefficient is only about 26% of the way from a random minimum to a lattice maximum. However, there are very few connected networks of $n = 4941$ and $k = 2.67$ for which $C > 0.08$, and those that do have much larger mean path lengths. Thus, in this case, SWI indicates that this network falls short of an impossible ideal, while ω' more usefully indicates that it comes close to the most small-worldly structure that is mathematically possible.

3 Discussion

Each of the indices of small worldliness considered here—Q, ω', and SWI—have advantages and drawbacks that should be considered when selecting a measure.

The primary advantage of Q is its simplicity, however, this advantage is outweighed by two disadvantages illustrated by the findings above. First, its use of a single reference network for both the clustering coefficient and mean path length is inconsistent with the WS conception of small-world structure. Second, its scale depends on n and k, making cross-network comparisons impossible. For these reasons and consistent with Telesford et al. (2011), although Q has been the most widely adopted index of small worldliness, alternative indices may be preferable.

Both ω' and SWI aim to overcome these drawbacks. They more appropriately compare the clustering coefficient to a lattice-based reference, but the mean path length to a random graph-based reference. They also offer a fixed scale (0 to 1 for ω' and SWI, but -1 to 1 for ω) that facilitates comparison of networks with

https://doi.org/10.1017/nws.2017.5 Published online by Cambridge University Press

different characteristics. The key difference between these two measures lies in their normalization schemes: ω' focuses on the extent to which the observed network's clustering coefficient and mean path length approach the values in target reference graphs, while SWI focuses on where a network's clustering coefficient and mean path length fall within the lattice-to-random range that defines the WS conception of small worldliness. This difference has implications for their interpretation. For a given set of constraints imposed on the reference graphs, there exists a network that achieves $\omega' = 1$, which represents a structure where clustering and connectedness are balanced. Thus, ω' captures the extent to which an observed network approaches this mathematically possible ideal. In contrast, $SWI = 1$ only when a network's clustering coefficient matches a corresponding lattice and its mean path length matches a corresponding random graph. However, such a graph does not actually exist. Thus, SWI captures the extent to which an observed network approaches this conceptual, but mathematically impossible, ideal.

These differences in normalization schemes point to these indices' drawbacks, which are illustrated in the real-world network examples in Table 2. The metabolic network illustrates that ω and ω' risk false-positives by not incorporating two key pieces of information: L_l or C_r. In contrast, the power grid network illustrates that SWI risks false-negatives by adopting a particularly stringent criteria for small worldliness. The choice between ω and SWI as an index of small worldliness should consider their differing interpretations and their tolerance for false-positives versus false-negatives.

This investigation has two notable limitations, which identify directions for future research that will help clarify this recommendation. First, the analyses above focus on a one-dimensional ring lattice and Erdös–Rényi random graph that preserve the observed network's n and k as the reference graphs. However, it will be important to explore these indices' behavior when computed using other types of reference graphs. Second, the performance of the indices was evaluated against the conceptual definition of small worldliness proposed by WS. Future studies should also explore their relative performance as predictors of phenomena hypothesized to be associated with small worldliness.

4 Materials and methods

Clustering coefficient. The clustering coefficient, C, measures the extent of clustering in a graph, but it has been measured in two different ways. WS measured it as

$$C_{ws} = \frac{1}{n}\sum_{i=1}^{n} \frac{\text{number of pairs of neighbors of } i \text{ that are connected}}{\text{number of pairs of neighbors of } i} \tag{10}$$

which conceptually captures the extent to which one's friends are also friends with one another. Others (Newman, 2010) have measures the clustering coefficient as

$$C_{\Delta} = \frac{3 \times \text{number of triangles}}{\text{number of connected triples}} \tag{11}$$

which is based on transitivity and captures the frequency with which potentially closed triples are closed. These two versions are typically in agreement, but in some

https://doi.org/10.1017/nws.2017.5 Published online by Cambridge University Press

cases can differ. Because C_{ws} was originally used in defining small-world networks, it is used in all analyses presented here.

Ring lattice construction. To construct reference ring lattices with given n and k, n nodes are first arranged in a circle at equal intervals. Next, a node with the lowest current degree (in the first round, any of the nodes, since all nodes begin with zero degree) is selected at random, and an edge is created between it and a random nearest neighbor in euclidean space to which it is not already linked. This process is repeated until the target k is achieved. This yields a ring lattice in which each node is connected to approximately k of its neighbors. When k is not an integer, each node in the resulting ring lattice is connected to either $\lceil k \rceil$ or $\lfloor k \rfloor$ of its neighbors.

References

Aldrich, H. E., & Kim, P. H. (2007). Small worlds, infinite possibilities? How social networks affect entrepreneurial team formation and search. *Strategic Entrepreneurship Journal*, **1**(1–2), 147–165.

Bagler, G. (2008.). Analysis of the airport network of India as a complex weighted network. *Physica A*, **387**(12), 2972–2980.

Bassett, D. S., et al. (2008). Hierarchical organization of human cortical networks in health and schizophrenia. *Journal of Neuroscience*, **28**(37), 9239–9248. doi: 10.1523/jneurosci.1929–08.2008.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, **10**(3), 186–198. doi: 10.1038/nrn2575.

Cardillo, A., Scellato, S., Latora, V., & Porta, S. (2006). Structural properties of planar graphs of urban street patterns. *Physica A*, **73**, 066107.

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.

Gallos, L. K., Makse, H. A., & Sigman, M. (2012). A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks. *Proceedings of the National Academy of Sciences*, **109**(8), 2825–2830.

Grannis, R. (2010). Six degrees of "who cares?". *American Journal of Sociology*, **115**(4), 991–1017.

Guimera, R., Danon, L., Diaz-Guilera, A., Giralt, F., & Arenas, A. (2003). Self-similar community structure in a network of human interactions. *Physical Review E*, **68**, 065103.

Guimera, R., Mossa, S., Turtschi, A., & Amaral, L. A. N. (2005). The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Procedings of the National Academy of Sciences*, **102**(22), 7794–7799.

Guye, M., Bettus, G., Bartolomei, F., & Cozzone, P. (2010). Graph theoretical analysis of structural and functional connectivity MRI in normal and pathological brain networks. *MAGMA*, **23**(5–6), 409–421. doi: 10.1007/s10334-010-0205-z.

Han, D. D., Qian, J. H, and Liu, J. G. (2009). Network topology and correlation features affiliated with European airline companies. *Physica A*, **388**, 71–81.

Humphries, M. D., & Gurney, K. (2008). Network 'small-world-ness': A quantitative method for determining canonical network equivalence. *PLoS ONE*, **3**, e0002051. doi:10.1371/journal.pone.0002051.

Humphries, M. D., Gurney, K., & Prescott, T. J. (2006). The brainstem reticular formation is a small-world, not scale-free, network. *Proceedings of the Royal Society B*, **273**(1585), 503–511. doi:10.1098/rspb.2005.3354.

https://doi.org/10.1017/nws.2017.5 Published online by Cambridge University Press

Kogut, B., & Walker, G. (2001). The small world of Germany and the durability of national networks. *American Sociological Review*, **66**(3), 317–335.

Lawlor, J. A., & Neal, Z. P. (2016). Networked community change: Understanding community systems change through the lens of social network analysis. *American Journal of Community Psychology*, **57**(3–4), 426–436.

Liu, Y., et al. (2008). Disrupted small-world networks in schizophrenia. *Brain*, **131**(4), 945–961. doi: 10.1093/brain/awn018.

Lusseau, D. (2003). The emergent properties of a dolphin social network. *Proceedings of the Royal Society of London B (Suppl.)*, **270**, S186–S188.

Masucci, A. P., Smith, D., Crooks, A., & Batty, M. (2009). Random planar graphs and the London street network. *European Physics Journal B*, **71**(2), 259–271.

Milgram, S. (1967). The small world problem. *Psychology Today*, **1**(1), 61–67.

Neal, Z. P. (2014). The devil is in the details: Differences in air traffic networks by scale, species, and season. *Social Networks*, **38**, 63–73. doi: 10.1016/j.socnet.2014.03.003.

Neal, Z. P. (2015). Making big communities small: Using network science to understand the ecological and behavioral requirements for community social capital. *American Journal of Community Psychology*, **55**(3–4), 369–380. doi: 10.1007/s10464-015-9720-4.

Neal, Z. P., Neal, J. W., Lawlor, J. A., & Mills, K. J. (2015). Small worlds or worlds apart? Using network theory to understand the research-practice gap. *Psychosocial Interventions*, **24**(3), 177–184.

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, **98**(2), 404–409.

Newman, M. E. J. (2010). *Networks: An Introduction*. New York: Oxford University Press.

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.

Sen, P., Dasgupa, S., Chatterjee, A., Sreeram, P. A., Mukherjee, G., & Manna, S. S. (2003). Small-world properties of the Indian rail network. *Physica A*, **67**, 036106.

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), 6.

Sporns, O., & Zwi, J. D. (2004). The small world of the cerebral cortex. *Neuroinformatics*, **2**(2), 145–162.

Telesford, Q. K., Joyce, K. E., Hayasaka, S., Burdette, J. H., & Laurenti, P. J. (2011). The ubiquity of small-world networks. *Brain Connectivity*, **1**(5), 367–375. doi: 10.1089/brain.2011.0038.

Uzzi, B., & Spiro, J. (2005). Collaboration and creativity: The small world problem. *American Journal of Sociology*, **111**(2), 447–504. doi: 10.1086/432782.

Wagner, A., & Fell, D. A. (2001). The small world inside large metabolic networks. *Proceedings of the Royal Society B*, **268**(1478), 1803–1810.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature*, **393**, 440–442. doi :10.1038/30918.

https://doi.org/10.1017/nws.2017.5 Published online by Cambridge University Press