2. Network and Nodal Indices. Measures of Complexity and Redundancy: A Review

Nigel Waters

PROLOGUE

In preparing this chapter, I was constantly amazed about how little some network researchers knew about what was going on in other disciplines where network analysis was thriving, and even in their own discipline where the topic had thrived at different times achieving various levels of popularity. Perhaps most disconcerting of all was that I realized that these observations also applied to my own knowledge of the field. Writing this chapter has helped me address some of my own deficiencies. Hopefully, it will be of use to others too.

2.1 INTRODUCTION

In 1973 I completed an MA thesis: ‘The Growth of the Southern Ontario Railway System – A Network Analysis’ (Waters 1973). This thesis drew heavily on a substantial body of literature within geography and spatial science that was devoted to network analysis. One impetus for this research had been the work of Karel Kansky (1963) whose own doctoral dissertation had been concerned with measures of nodal importance and network complexity within a transportation network. These were believed to be associated with urban growth. If this were the case, then socio-economic variables might be used to simulate the growth of such transportation networks as the road and rail system of a country.

Kansky’s work spawned a decade of increasingly sophisticated research into transportation and related networks. Some of this was summarized in Haggett and Chorley’s (1969) book *Network Analysis in Geography*, which attempted to forge a link between physical and human geography. The authors were prescient in their realization that network analysis might be applied successfully to many disciplines, but their attempt to unify their own discipline through this mechanism was fruitless. Even today the discipline of geography remains hopelessly divided (Waters 2003), and recently some have even called for its permanent division (Thrift 2002) into physical and human geography.

Although there was an attempt to move beyond what was mistakenly seen as an arid concern with static, structural concerns (for example, Black 1969, 1971; Taaffe and Gauthier 1973; and even Kansky 1963 had produced time-stamped, simulation growth models as noted above), spatialism had ceased to be the paradigm of the moment. Geography, described by Taylor (see Waters 1993a) as the Latin America of disciplines because it moved rapidly, in revolutionary fashion, from one paradigm (school-of-thought, fad, ruling clique?) to another, now moved on to a concern with relevance and structuralism and to Marxism, and then to a host of 'post-thises' and 'post-thats' (for example, post-Marxism and post-structuralism, Waters 1993a). Paradoxically, the interest in structuralism did not extend to the structure of networks when reduced to graphs or to social network analysis. Within the disciplines of geography and regional science there has always been a small chorus that sang from the spatial science song sheet (Batty and Longley 2003), although for many years this song sheet did not include network analysis to any substantial degree. Gradually, as Geographic Information Systems (GIS) software grew more powerful in the 1980s the chorus grew louder and research on spatial analysis and indeed spatial science and a Geographic Information *Science* (Goodchild 1992) returned to the mainstream.

Network analysis, however, remained somewhat of a backwater. There were many developments in shortest path modelling, since this was a fundamental building block for location-allocation modelling (at least the discrete space versions) and for the four-step model, the *sine qua non* of transportation planning (see, for example, www.caliper.com). Moreover, the coding and characterization of network topological structures within GIS databases was also an active research frontier, but the *structural* analysis of networks *per se* was essentially abandoned.

2.2 OTHER DISCIPLINES AND NETWORK ANALYSIS

This lack of interest was not true of other disciplines.

Most popular texts on networks note that mathematicians have been interested in the topic since Euler laid the foundations of graph theory in

1736 when he proved that there was no path that allowed the seven bridges of Konigsberg to be crossed once and once only because more than two nodes had an odd number of links (Barabási 2002, pp. 9–13). More recently the work of Erdos and Renyi (1959, cited in Barabási 2002; Schintler et al. 2005a) on the theory of random graphs is usually considered seminal by those describing the history of the field. A fundamental finding in the Erdos–Renyi paper was that, as the average number of links per node rose above 1, the fraction of nodes that were connected into the single largest component jumped dramatically. Unfortunately, networks that exist in the real world are not random graphs. Usually the local linkages are much stronger and the more distant linkages much weaker than they are in a random graph. Friends in social networks and links in transportation systems, other things being equal, tend to favour the local over the global, the shorter distances over the longer.

2.3 RESEARCH BY SOCIOLOGISTS

Sociologists have maintained an ongoing interest in social network analysis and social structure analysis for four decades now. For many, this interest stemmed from the work of Harvard-based sociologist, Stanley Milgram (Milgram 1967; and see Travers and Milgram 1969 for technical details; also Batty 2001; Waters 2001; Barabási 2003; Watts 2003). Milgram had a gift for designing highly original, empirical experiments (for a discussion of Milgram's experiments on subservience to authority, see Blass 2003).

Milgram was intrigued by the notion that we are all linked to each other through our social networks. His experiment involved determining how many steps it would take for someone living in Boston to get a letter to someone living in Nebraska without using the direct postal system. The letter could be sent only to individuals that the subject knew on a first-name basis. The subject selected these individuals on the basis that they would get the letter closer to its predetermined destination. Travers and Milgram (1969) in the subsequent more explicit discussion of their methodology noted that many of the subjects did not complete their chains, but for those that did the average was about six steps.

Although Milgram's experiment has been questioned in recent literature (Kleinfeld 2002; see also http://www.columbia.edu/itc/sociology/watts//w3233/client_edit/big_world.html), it has given rise to the popular concept that everyone on the planet is linked by just six degrees of separation (that is, where a degree is one link in the chain of individuals). Despite widespread admiration for Milgram's experimental skills, the task was probably too easy in some respects since it gave the target's name and address. However, there was no optimizing, teleological algorithm to guide the subjects on the most profitable shortest path. So, even if we are all separated by six degrees (or links), picking the right links between hubs in our social networks to move a message along an optimal chain is a difficult task and hence the huge investments of intellectual expertise in designing fast, web search algorithms and shortest path algorithms (Roush 2004).

This difficulty in determining just 'who you're gonna call' in order to get the message through was illustrated when a reporter, Mike Wilson, at the St. Petersburg Times in Florida ran an e-mail-based experiment to see if a target by the name of Tara could be found by e-mail by a small number of subjects. The only information that was provided was as follows: 'Tara's first name and a few biographical details: approximate age, town in which she grew up, marital status, the U.S. region she lives in, the industry in which she works and some family background' (http://www.sptimes.com/News/063000//Floridian/The_search_for_the_Ta.shtml). The search proved to be fruitless (http://www.sptimes.com/News/081300/Floridian/Pondering_six_degrees.shtml). Tara was never found and, in the subsequent article, her name was revealed. Kleinfeld felt that this was evidence that the world is much bigger than Milgram first suggested (http://www.columbia.edu/itc/sociology/watts//w3233/client_edit/big_world.html).

Perhaps a better six-degrees-of-separation experiment in the Internet age might be to investigate e-mail address books. Acquiring sample sizes that would be large enough and unbiased might prove to be a challenge (Lakhina et al. 2003). It would, however, provide more definitive answers to how many links the average person has and how tightly knit the networks are. Watts and his supervisor, Strogatz (Watts and Strogatz 1998; Watts 1999), developed an index to measure this interlinkage that they called the 'clustering coefficient'. This was a measure of the observed number of first-order links over the maximum possible number of first-order links. For four friends, the latter number would be six, which represents a situation where each of the four friends is a friend of the other three individuals and is so linked. If only four of the possible six links existed then there would be a clustering coefficient of 0.66. Barabási (2002, p. 249) notes that the same measure was used in the sociological literature where it was referred to as the 'fraction of transitive triplets' (Wasserman and Faust 1994). Kansky (1963) had, of course, used a similar index, the gamma index, as a measure of *network* connectivity. The gamma index for non-planar graphs measures the number of observed links to the maximum possible number of links and is available in non-planar and planar versions (ibid.). If it is limited to the immediate neighbourhood of first-order links around a given point, it is equivalent to Watts and Strogatz's clustering coefficient.

Such an observation invites comparison to the order statistics developed in point pattern analysis for nearest neighbour distances. The original nearest neighbour statistics were just that, a comparison of the observed mean distance to each point's nearest neighbour with the expected mean for a Poisson (or some other theoretical) distribution. Subsequent researchers extended this to the second-, third-, fourth-, fifth- and sixth-order nearest neighbours in order to capture more fully the complexity and subtlety of the pattern (Jones 1971). Similarly, the clustering coefficient might also be calculated between the extremes of the Watts–Strogatz measure and the gamma index (Kansky 1963; and see discussion below). Such an approach is also in sympathy with recent attempts to develop an arsenal of Local Indicators of Spatial Association (LISA) statistics (Anselin 1998), as opposed to the preference for global indices of earlier decades. It also resonates with attempts to develop measures of autocorrelation within the network structure (Doreian 1989a, 1989b; and, in geography, see Arthur 2002, and the references therein).

It is intriguing to speculate on the origin of this idea of a 'small world', but one early reference has been attributed to Jane Jacobs (1961, pp. 145–51). In her classic text, *The Death and Life of Great American Cities*, Jacobs notes how she used to play a game with her sister after they moved to New York. The game was called Messages and the object was to specify the shortest set of links that would allow a message to be passed by word of mouth between 'two wildly dissimilar individuals – say a headhunter in the Solomon Islands and a cobbler in Rock Island, Illinois'. The winner was the one with the shortest, possible chain. Jacobs's description of her example chain is illuminating (ibid., p. 145):

> The headhunter would speak to the headman of his village, who would speak to the trader who came to buy copra, who would speak to the Australian patrol officer when he came through, who would tell the man who was next slated to go to Melbourne on leave, and so on. Down at the other end, the cobbler would hear from his priest, who got it from the mayor, who got it from the state senator, who got it from the governor, and so on.

Two features of real-world networks are identified here. First the importance of *local connectivity*; and, second, the importance of *hierarchical linkages* which connect local networks to each other (note the use of the expression 'down at the other end'). We will see this theme emerge in later (and earlier) work in various fields. Jacobs had to admit that connecting the local networks was the tricky part and so they invented a character called Mrs Roosevelt who had ties to a vast number of individuals (ibid., p. 145): 'Mrs Roosevelt made it possible to skip whole chains of intermediate connexions. She knew the most unlikely people. The world shrank remarkably'.

We will see that sociologists writing over a decade later emphasized the importance of such individuals who have loose ties to large numbers of local networks. Jacobs, herself notes the importance of such people in binding the fabric of a community together (ibid., p. 145): ‘A city district requires a small quota of its own Mrs Roosevelts – people who know unlikely people, and therefore eliminate the necessity for long lines of communication (which in real life would not occur at all)’.

Interestingly, Barabási (2002, p. 253) states that, just prior to Jacobs’s publication of her book, Rosenthal had discussed the acquaintances and contacts of Mr Roosevelt in his Master’s thesis (Rosenthal 1960). Even today, such transnational links continue to be the focus of much sociological research (Waldinger and Fitzgerald 2004, p. 1177): ‘social scientists are looking for new ways to think about the connections between ‘here’ and ‘there,’ as evidenced by the interest in the many things called *transnational*’.

The complexity of such urban social networks has recently been addressed by Oldfield (2004) to determine the degree of integration in the racially desegregated area of Delft South, a low-income neighbourhood in Cape Town, South Africa.

Recently, interest in networks among social scientists has even spawned new journals such as *Global Networks*, which was launched in 2001. As one set of contributors to the inaugural volume noted (Dicken et al. 2001, p. 89): ‘This paper advocates a network methodology as a potential framework . . . Such a methodology requires us to identify actors in networks, their ongoing relations and the structural outcomes of these relations’.

In 1973, Mark Granovetter, after an initial rejection, finally had his paper, ‘The strength of weak ties’, published in the *American Journal of Sociology*. It quickly became a citation classic. Granovetter’s paper proposed that, when it comes to getting a job (and a host of other interactions of a socio-economic nature), it is our weak links that are most important. These weak links, as opposed to the strong links that knit together our family and closest friends, are the links that are important in helping us find jobs, get information and, as we shall see below, adopt new innovations. It is the weak links that allow us to link our own cluster of friends to other groups. Quite simply put, they are the Mrs Roosevelts of the world. They allow us to establish a hierarchy within the network, a hierarchy that may have many levels. In certain transportation networks, such as airline networks, these are the connectors between hubs (Taaffe et al. 1996). It is these long-distance links that can turn our world into a small world. It produces the ‘scale-free networks’ that Watts and Strogatz (1998) showed were so different from the random networks of Erdos and Renyi (1959; cited in Watts 2003).

The Erdos–Renyi random graphs and the hierarchical small-world graphs (now known as scale-free networks, which are so common in the real world) are both examples of what Kansky and earlier researchers chose to call 'non-planar graphs'. That is to say, a non-planar graph is one in which links can cross without a node being located at the crossing point. Although roads and railway networks can have links that cross each other without a node being located at the junction (for example, bridges) they are probably best described as planar networks (Waters 1973). Barabási (2002, p. 69; for a discussion of the distinction between scale-free and exponential networks, see also Albert et al. 2000) recognizes this difference between, on the one hand, the scale-free networks that are now shown to exist on the Internet and for airline systems, and, on the other hand, road networks, where there 'are no cities served by hundreds of highways' (Barabási 2002).

It is interesting to speculate on just what the degree distribution would be for a planar road network. This does not appear to have been done within the literature to date. What would its degree distribution look like? Would it follow a very shallow power law with a remarkably gentle slope? Would such a network *have* to have a huge diameter, many degrees of separation, an incredibly large average graph distance? Is it distinctively different from the scale-free networks that are now being actively examined. Some of these topics are discussed below along with a definition of the terms that have been introduced. It is possible that these questions may pose an entirely new research agenda.

Work by sociologists continues unabated. Major recent contributions of texts include the work of de Nooy et al. (2005) on the use of the Pajek software for exploratory social network analysis, and Carrington et al.'s (2005) on models and methods, which includes Everett and Borgatti's (2005) work on extending centrality concepts and measures.

Before leaving the work of the sociologists, it is important to note that they have developed a sizeable arsenal of computer programs for analysing social networks. These programs are reviewed in Wasserman and Faust (1994, Appendix A). More recent and comprehensive reviews may be found on the Internet at: http://www.sfu.ca/~insna/INSNA/soft_inf.html, while a bibliography of social network methodologies may be located at: http://faculty.ucr.edu/~hanneman/SOC157/TEXT/Bibliography.html.

2.4 COMPUTER SCIENTISTS

Computer scientists in the 1980s and early 1990s turned to some of the more fundamental indices of network structure that Kansky (1963) had employed as measures of software reliability. These indices included the cyclomatic number. The latter is a measure of network complexity that is a count of the number of circuits in the network given by the formula:

$$L - N + G, \tag{2.1}$$

where L is the number of links, N is the number of nodes, and G is the number of separate subgraphs; if the network is nowhere disconnected, this number is always 1. The higher the cyclomatic number, the more circuits there are and the greater the network complexity. In programming, the circuits are loops in the code and the greater the number of loops the greater is the chance that the software will fail in a given length of time (Kitchenham 1990). This correlation between program complexity, as measured by the cyclomatic number, and the mean-time-to-fail rate is of particular concern in those situations where the software is used to operate transportation systems (Waters 1993b).

2.5 PHYSICISTS AND MATHEMATICIANS

The contributions of mathematicians and physicists such as Watts and Barabási have been seminal since 1998 when Watts (and Strogatz) first provided mathematical characterizations of the small-world phenomenon from a network viewpoint. These contributions have been discussed extensively above and will again be commented upon below.

Mathematicians have provided researchers with up-to-date resources and a compendium of new fast algorithms (Gross and Yellen 2004).

2.6 DIFFUSION OF INNOVATIONS

Innovation diffusions result, in part, from social networks. To adopt, a user has to be convinced of the value of the new product. Rapid diffusion depends on innovators who are willing to act as product champions. If this occurs, an innovation can, after a slow start, be adopted rapidly. It has long been known that the cumulative frequency distribution of the adoption curve can be expressed as a logistic function (Everett 1962, 2003; Watts 1999; Longley et al. 2001; see also Moore 1991, who describes the move from slow to rapid adoption as 'crossing the chasm', while Gladwell 2000 refers to it as the 'tipping point').

The reason for this logistic function is due to both the local connectivity of the network and the hierarchical aspects of the system. Much of the early work on innovation diffusion was performed by rural sociologists and many of the innovations studied were farming innovations. The difficulty of establishing a hierarchical network among such a disparate population was long realized in the United Kingdom, and one method of reaching everywhere quickly was through the radio and in particular radio soap operas such as *The Archers*, which was targeted at the farming community. The Archers were the Mrs Roosevelts of the UK farming world.

2.7 GEOGRAPHERS AND REGIONAL SCIENTISTS

Perhaps one of the most exciting developments in network analysis among regional scientists and related disciplines such as transportation science and geography is the work of Anna Nagurney (http://www-unix.oit.umass.edu/ /~nagurney) and her colleagues at the Virtual Center for Supernetworks (http://supernet.som.umass.edu/) at the Isenberg School of Management at the University of Massachusetts at Amherst. Despite the huge interest in the topic of supernetworks (see Nagurney and Dong 2002, the first in a series on New Dimensions in Networks that now includes a second and a third volume: Nagurney 2003, and Lee 2004), this work will not be reviewed here. Rather this chapter will concentrate on the narrow topic of the topological characteristics of transportation and similar networks. If network topology is of such interest to sociologists, mathematicians, biologists and computer scientists, perhaps it is time that geographers and regional scientists also renewed their concern.

ESRI, the leading GIS software manufacturer, has been concerned over the development of new functionality and new data structures for networks, and these have been implemented in the most recent release of its ArcGIS software (version 9) together with its associated Network Analyst extension (see ArcNews Online Spring 2005 edition at http://www.esri.com/news/ /arcnews/spring05articles/advanced-modeling.html and http://www.esri.com/ /news/arcnews/spring05articles/transportation-model.pdf). Nevertheless, the type of analysis conducted by researchers investigating the phenomenon of scale-free networks could not easily be carried out with ESRI's software. Thus GIS researchers wishing to perform this type of analysis will have to continue to use specialized software or wait for new extensions and scripts to be written that will satisfy GIS users.

Indeed, geographers have recently published a number of articles that revisit the topic of network topology and have not only given new life to the old measures of network structure but have also introduced new indices of network complexity and nodal importance.

In the former category is the work of O'Hagan and Green (2004). This research examined corporate directors for 750 firms for the years 1976, 1986, and 1996 (these firms included the Fortune 500 firms in the US and the top 200 industrial and service firms, plus the top 50 financial institutions in Canada as identified from the *Financial Post*). This information was recorded in an interlocking matrix of directorships for 195 North American cities (151 US and 44 Canadian cities). The matrix was then analysed to determine the centrality of these city nodes using the simplest of measures: namely, the degree or valency or number of links possessed by each node. This is then compared with the number of links in the network. Using the former number as the numerator and the latter as the denominator provides the measure of centrality. In the US this measure, when applied to New York, tended to decline over the 20-year study period, while Toronto retained its dominance over the Canadian corporate network.

There really are a host of indices that measure centrality more effectively (Tangmunarunkit et al. 2001) but these are not discussed in the article. Thus we can establish that many authors not only do not know what is going on in network analysis in fields other than their own but also are not aware of the history of network analysis in their own disciplines.

O'Hagan and Green also make use of Nystuen and Dacey's methodology (1961; also Taaffe and Gauthier 1973, pp. 149–58) for establishing regional hierarchies within the corporate networks. A city is thus determined to be independent if its largest flow is to a smaller city. The hierarchy is enforced by requiring the property of transitivity to be maintained among triads of cities and by not allowing a city to be subordinate to any of its own subordinates. The methodology allows O'Hagan and Green to establish that there are a number of regions within the US corporate hierarchy but only one, centred on Toronto, in the Canadian network. This work emphasizes that, within these corporate networks, as in many others, there are both regional and hierarchical effects. As has been shown in other research venues, this has important implications for the diffusion of innovations and disease transmission, as well as the knowledge transfer between the Fortune 500 companies with which O'Hagan and Green were concerned.

Recently, network models have been employed by wildlife geographers and biologists to study the destruction of animal habitats. Researchers identify habitat patches using software such as Fragstats, which is now available as an extension to ESRI's ArcView software (Longley et al. 2001, p. 313). Perhaps the best example of such work is Bunn et al.'s (2000) study of landscape connectivity, which is one of a very small number of studies that examine the impact of node and link loss on the structure of the fragmented landscape. They do not, however, devise more than a subjective way of handling the diameter of the reduced network (ibid., p. 270). This concern will be revisited below.

2.8 BRINGING THE DISCIPLINES TOGETHER

Much recent work has led to a merging of disciplinary perspectives. Thus physicists are now studying explicitly spatial networks, while geographers, regional scientists and economists are integrating the ideas of scale-free networks developed by the physicists into their studies of transportation and information networks (much of this work, updated on a monthly basis, can be found archived at the Econophysics Forum website: http://www.unifr.ch/ /econophysics).

Important among the work of the physicists in linking spatial graphs with the spatial or topological relational graphs is Watts's first book (1999, pp. 91–100, 127–37). Watts investigated the properties of spatial graphs as well as relational graphs, showing that in the former there was a preferential attachment over short distances. As the distance cut-off point was increased in the spatial graphs, they began to exhibit the same sort of properties shown by the relational graphs. In a study of the crossover from scale-free to spatial networks, Barthelemy (2003) has developed a model that incorporates both preferential attachment and distance selection, which is characterized by some typical finite interaction range. The optimal network that minimizes both the total length of the edges and the diameter is shown to lie in-between the scale-free and spatial networks. It is interesting to speculate where on the continuum this point might lie for different modes, such as air, rail, road and pipeline networks. Obviously, the interaction range will affect the location of this point, and investigations into its location on the continuum will likely be the focus of future research.

The very latest research (June 2005) at the Econophysics website has shown that networks that are scale free may also exhibit self-similarity and thus fractal-like properties (Song et al. 2005). The journal *Nature* continues to provide a rich vein of articles on network topology, including the recent contribution by Palla et al. (2005), which explores the overlapping community structure of complex networks, both natural and societal. Biological examples offered by Palla et al. include protein structures, but their approach might easily be applied to the habitat fragmentation case studies explored by Bunn et al. (2000), discussed above. Thus Palla et al.'s methodology could be used to explore the degree of community overlap *for different species* on the basis of a network representation of their habitat fragmentation. In addition, Farmer et al. (2005) argue that economics may be the next physical science and that 'remarkable regularities in economic data may suggest parts of social order that can be usefully incorporated into … the conceptual structure of physics'.

Lakhina et al. (2002) have examined the Internet's physical structure (a topic that has become a subdiscipline in its own right, with regular meetings of the Association for Computing Machinery (ACM) Special Interest Group on Data Communication). Their research shows that router density varies widely over different economic regions although within economically homogeneous regions it is related to population density. Also, router connection is 75 to 95 per cent dependent on geographical distance. Finally, autonomous systems within the Internet structure that are above a certain size show maximum geographical dispersal. These findings would imply that a location-allocation modelling approach where population density was a demand variable might be used successfully to plan the next generation in Internet infrastructure. The importance of distance and population again reflects the seminal work of Black (1969, 1971) on simulating railroad growth, although the influence of angular parameter (that is, deviation from the main line) seems unimportant in the structure of the Internet.

For geographers, Batty's (2001) recent editorial has been one of the clarion calls to integrate small world research into geographic analysis especially of the urban economy (see also Andersson et al. 2003). Batty's own earlier work on fractals may now be integrated into this research stream. Batty (http://www.casa.ucl.ac.uk/people/MikesPage.htm), commenting on the developments in economics and physics and presumably geography, has even suggested the eventual formation of a 'psychohistory' akin to the ideas presented in Isaac Asimov's *Foundation Trilogy* (1982) or, more prosaically, in Wilson's book *Consilience* (1998). More recently, Batty (2003) has expanded upon these ideas, arguing for the integration of the new developments from physics into a more robust Geographic Information Science. Batty (2004) has also advocated the integration of graph-theoretic and distance measures with Hillier and Hanson's (1984) space syntax methodology. This new direction has in turn been championed by physicists (Porta et al. 2005) in a new contribution that introduces new measures of multiple centrality assessment.

Perhaps most important of all with respect to attempts to integrate the work of physicists on networks into social and regional science has been the research of the George Mason School of Public Policy and their European collaborators. Recent papers that provide an integrated view of the topological and spatial views of networks include Gorman and Kulkarni (2004) and Schintler et al. (2005b).

The integration of geography and physics is again an idea that has been around for a while. Indeed, it was very much the focus of William Warntz's long and distinguished career (Warntz 1973).

2.9 OTHER MEASURES OF NETWORK COMPLEXITY, STRUCTURE AND REDUNDANCY: LINKING THE RESEARCH

Apart from some notable exceptions, there has been a curious tendency to use existing and quite dated measures of network structure. Regional scientists studying the Internet (for example, Wheeler and O'Kelly 1999) and geographers, such as O'Hagan and Green (2004) studying the corporate structure in North America, have relied on indices that were first advocated more than 40 years ago. Other researchers have ignored developments in their own disciplines and other fields of research that have used network analysis and cognate statistical techniques.

Work on the degree structure has been used as a measure of centrality in studies such as O'Hagan and Green (2004). However, there are more subtle measures that might well be more useful. These measures include the row sums of the shortest path matrix, and the rows sums of the powered binary connectivity matrix. Both have been in the literature for over 30 years (Taaffe and Gauthier 1973), although both have been modified slightly in recent work (Lee and Lee 1998; Wheeler and O'Kelly 1999). Watts and Strogatz (1998) and Barabási (2002) and his colleagues have introduced new ideas by comparing the degree distribution to mathematical distributions such as the power law. This allows observed networks to be compared to norms so that deviations can be determined and studied, and hopefully explained. This work is reminiscent of the early work on stream networks (discussed in Haggett and Chorley 1969) where laws of stream numbers were eventually found to be derivable from distributions of topologically distinct channel networks (TDCNs), where each TDCN is assumed to be equally likely.

As noted above, in the early work on point pattern analysis, nearest neighbour statistics were calculated for first nearest neighbours. Eventually, this work was extended to the first six regional nearest neighbours, and then Jones (1971) provided confidence intervals for random disturbances on regular point lattices by means of simulation experiments. Network analysis would benefit from having statistics for observed degree distributions for second and subsequent nearest neighbours in the network. These could be compared to random expectations and, perhaps more profitably, to expectations for power-law models. Simply stated, it would provide a greater understanding of the network structure beyond its nearest neighbours.

Power laws are found in urban geography with the rank-size rule for urban population sizes, in the distribution of oilfields (Lee and Wang 1986; Lee and Gill 1999) allowing exploration to be targeted at missing members, and in trip distributions. Research in the early 1970s linked linear cost-minimizing models from the spatial interaction tradition with the entropy-maximizing work of Wilson and his colleagues (Evans 1973; Wilson 1970). Nijkamp and Reggiani (1989) went further and were able to integrate spatial interaction and input–output models within a dynamic stochastic multi-objective framework. Their chapter produced some remarkable results showing that stochastic fluctuations tend to destabilize spatial interaction systems. This has implications for the modelling of many systems, including electricity power grids. These have traditionally been modelled using system dynamics approaches, and such models have tended to show not only how unstable these systems are but also how balancing loops can be introduced into the system to re-stabilize them. Thus Seel (2004; see also Watts 1999, p. 147 for a discussion of power grid networks) has shown how awareness of the cost of power can re-stabilize a deregulated power grid thus preventing a series of boom–bust price cycles.

Nijkamp and Reggiani (1989) link their work to fluctuations in biological models and this resonates with Barabási's attempts to link his degree distribution functions, for say the Internet, with those developed for models of cell structure. The former authors also demonstrate linkages with logit models that most regional scientists will be familiar with from their use in models of modal choice in the four-step, transportation planning model.

2.10 NODAL REDUNDANCY

In the early work on network structure, few researchers looked at nodal redundancy, although Reed (1970) was a notable exception. There are at least three reasons why nodal redundancy might be of concern. First, it is important to recognize that some networks are contracting or being rationalized. This, for example, was true of many rail networks in the 20th century, and it is also true for areas of habitat that can be represented as networks, as in the research of Bunn et al. (2000). Second, it is important to note that the Internet was designed to make use of redundancy in the network structure (Wheeler and O'Kelly 1999). The goal was to allow messages and information sent over the network to be split up into packets that could be re-routed if the system was busy or if a node went out of service. A third reason for needing to determine the importance that might be attached to the loss of a node within the network concerns security. From a defence viewpoint the lack of redundancy in Canadian transportation systems has been an ongoing concern (Lea and Waters 1996). Geographers and regional scientists have been involved increasingly in attempts to improve security, especially in North America (Cutter et al. 2003). Crampton (2004) states that the Directorate of the Information Analysis and Infrastructure Protection, a section of the Department of Homeland Security in the US, recently made a budget request for 'development and maintenance of a complete and accurate mapping of the Nation's critical infrastructure and key assets'. Mapping the infrastructure is only the beginning of security. A critical next step is to know which of the nodes in the network would have the greatest impact on the network structure if they were lost. It should, however, be noted that a recent report prepared by the Rand Institute in the US on the implications of publicly available geospatial information pays exclusive attention to the potential loss of critical *sites* and does not discuss the damage that the loss of a node *within a network* would occasion (Baker et al. 2004).

It might be possible to reverse the simulation methodology developed by Black (1971) based on discriminant analysis. Thus a discriminant function could be constructed to determine which nodes could be removed from the network while minimizing the loss of revenue to the network as a whole. The logic of such an approach can also be applied to the traditional measures of nodal importance discussed above: namely, the row sums of the shortest path matrix and the row sums of the sum of the powered binary connectivity matrix. Reed (1970) took the necessary next step when he successively removed nodes from the network and then calculated the average graph distance (AGD) for each of the remaining nodes in the network. This measure shows how the connectivity, in terms of the average shortest path, is affected for each and every node in the network. Reed then calculated the mean change in the absolute value of the AGD for the entire network. It was an interesting and pioneering (though not seminal) attempt to measure the loss of a node in the network.

There were, however, two problems with Reed's measure. The first was the use of the absolute value of the AGD. If certain peripheral nodes were lost from the network, it was possible, perhaps likely, that the AGD would decrease, making the network better connected. This, after all, was the rationalization for removing what were seen as redundant nodes in 20th century railway networks. However, using the absolute value of the mean change in AGD after the loss of a node would give an erroneous impression of its importance. The second difficulty with Reed's measure occurred when the network became disconnected and split into two or more subgraphs with the loss of one or more nodes. Without modification, Reed's measure becomes problematic and cannot be computed for the entire, albeit, disconnected network.

The solution to these difficulties was to use the real change in mean AGD for the network, whether it was a positive or a negative value. The value or worth of a node within the network structure could then be determined by a ranking of these changes. In addition, a penalty would have to be introduced for disconnecting the network. There is no shortest path between a node in one subgraph and a node in a second subgraph. One method for 'estimating' the length of such a path, even though arbitrary and subjective, is to set the path as being one link greater than the longest shortest path (that is, the diameter) in the new network. An alternative would be to insert some penalty related to the number of subgraphs produced, but this would be far less sensitive to the impact of the loss of a node.

It should be realized at this point that Reed's average graph distance is the same as Milgram's six degrees of separation for our social networks or Albert et al.'s (1999) 19 degrees of separation for the Internet.

If we modify Reed's Measure of nodal importance and use positive and negative changes and include some penalty measure for disconnecting the network, then this new measure, Reed's Measure Modified, can be used as a measure of nodal importance for all the situations that we have described in the chapter so far, that is, social networks, the Internet, financial and transportation networks, and habitat fragmentation, among others. We have written a Visual Basic Program that calculates this index for networks. To date, it has been applied to small, test networks and has shown results that indicate that the index correlates only moderately with the other traditional network indices. In other words, this index does measure something new and different. It is now our intention to apply it to other networks that have been examined using these traditional indices as measures of centrality. Others too have begun to work on this problem, and Junelius and Petersen's (2005) contribution measuring link and node vulnerability of the northern Swedish road network is noteworthy.

2.11 CONCLUSION

This chapter has provided a brief review of the history of the structural analysis of the topological properties of networks in a variety of fields including mathematics, geography and regional science, physics and biology, and sociology. It has also shown that, while this interest in network topology has never completely gone away, it was not a major concern for many of these disciplines in the 1970s, 1980s and early 1990s. In the late 1990s, this changed quite remarkably. The cause of this change was a renewed interest of mathematicians and physicists, stimulated by the work of Watts and Strogatz (1998), in the characteristics of real-world networks rather than the random structures studied by Erdos. This led to the derivation of new empirical laws of network structure.

A second reason for the increase of interest in network structure was the exponential growth of the Internet. Characterizing the structure of the Internet is important if we are to build better, faster and more effective search engines (Roush 2004). By adding more links to our websites, we can drastically reduce the average graph distance or the number of degrees of separation in the network.

Finally, a concern with security has meant an emphasis on the structure of networks that support the economy. Financial networks, transportation networks and the Internet all must have a measure of redundancy built into them (Lea and Waters 1996). If a node is attacked and removed in a network with redundant nodes and links, its traffic and services can be supported by other nodes and links within the network. Determining the degree of importance of such nodes requires new indices such as that developed by Reed, together with the modifications suggested in this chapter.

Establishing the same laws for these new measures, as Barabási (see Barabási 2002 and the references therein) and his colleagues and others have done for such traditional measures as the degree distribution, is the next research challenge. This, however, must be achieved using data sets that are not biased samples from the Internet and other sources, as argued by Lakhina et al. (2003). Linking researchers in the disparate disciplines that have a strong interest and track record in network analysis is also a major challenge. This chapter has attempted to identify some of these links.

REFERENCES

Albert, R., H. Jeong and A.-L. Barabási (1999), 'Diameter of the World Wide Web', *Nature*, **401**, 130–31.

Albert, R., H. Jeong and A.-L. Barabási (2000), 'Attack and error tolerance of complex networks', *Nature*, **406**, 378–82.

Andersson, C., A. Hellervik, K. Lidgren, A. Hagson and J. Tornberg (2003), *The Urban Economy as a Scale-Free Network*, http://www.unifr.ch/econophysics/ March.

Anselin, L. (1998), 'Exploratory spatial data analysis in a geocomputational environment', in P.A. Longley, S.M. Brooks, R. McDonnell and W. Macmillan (eds), *Geocomputation: A Primer*, New York, NY, USA: Wiley.

Arthur, R. (2002), 'Spatial and temporal analysis of traffic collisions', Unpublished PhD Dissertation, Department of Geography, University of Calgary, Calgary, Alberta.

Asimov, I. (1982), *The Foundation Trilogy*, New York, NY, USA: Doubleday.

Baker, J.C., B.E. Lachman, D.R. Frelinger, K.M. O'Connell, A.C.Hou, M.S. Tseng, D. Orletsky and C. Yost (2004), *Mapping the Risks: Assessing the Homeland Security Implications of Publicly Available Geospatial Information*, RAND National Defense Research Institute, Santa Monica, CA, http://www.rand.org/ /publications/MG/MG142/MG142.

Barabási, A.-L. (2002), *Linked: How Everything is Connected to Everything Else and What It Means for Business, Science and Everyday Life*, New York, NY, USA: Perseus.

Barthelemy, M. (2003), ‘Crossover from scale-free to spatial networks’, *Europhysics Letters*, **63** (6): 915–21.

Batty, M. (2001), ‘Editorial: cities as small worlds’, *Environment and Planning B: Planning and Design*, **28** (5): 637–8.

Batty, M. (2003), *Network Geography: Relations, Interactions, Scaling, and Spatial Processes in GIS*, Working Paper 63, Centre for Advanced Spatial Analysis, University College London http://www.casa.ucl.ac.uk/working_papers/paper63.pdf.

Batty, M. (2004), *Distance in Space Syntax*, Working Paper 80, Centre for Advanced Spatial Analysis, University College London http://www.casa.ucl.ac.uk//working_papers/paper80.pdf.

Batty, M. and P. Longley (2003), *Advanced Spatial Analysis: The CASA Book of GIS*, Redlands, CA, USA: ESRI Press.

Black, W.R. (1969), ‘The generation of transportation networks: their growth and structure’, PhD Dissertation, Department of Geography, University of Iowa, Iowa City, Iowa.

Black, W.R. (1971), ‘An iterative model for generating transportation networks’, *Geographical Analysis*, **3**, 283–8.

Blass, T. (2002), ‘The man who shocked the world’, *Psychology Today*, **35** (2), 68–75.

Bunn, A.G., D.L. Urban and T.H. Keitt (2000), ‘Landscape connectivity: a conservation application of graph theory’, *Journal of Environmental Management*, **59** (4), 265–78.

Carrington, P.J., J. Scott and S. Wasserman (eds) (2005), *Models and Methods in Social Network Analysis*, Cambridge, UK: Cambridge University Press.

Crampton, J.W. (2004), ‘Rethinking GIS and (homeland) security’, *GeoWorld*, **17** (4), http://www.geoplace.com/gw/2004/0404/0404cgis.asp.

Cutter, S.L., D.B. Richardson and T.J. Wilbanks (eds) (2003), *The Geographical Dimensions of Terrorism*, New York, NY, USA: Routledge.

Dicken, P., P.F. Kelly, K. Olds and H.W.-C. Yeung (2001), ‘Chains and networks, territories and scales: towards a relational framework for analysing the global economy’, *Global Networks*, **1**, 89–112.

Doreian, P. (1989a), ‘Network autocorrelation models: problems and prospects’, in D.A. Griffith (ed.), *Spatial Statistics: Past, Present and Future*, Ann Arbor, MI, USA: Institute of Mathematical Geography, pp. 369–89.

Doreian, P. (1989b), ‘Two regimes of network effects autocorrelation’, in M. Kochen (ed.), *The Small World*, Norwood, NJ, USA: Ablex, pp. 280–95.

Evans, S. (1973), ‘A relationship between the gravity model for trip distribution and the transportation problem in linear programming’, *Transportation Research*, **7**, 39–61.

Everett, M.G. and S.P. Borgatti (2005), ‘Extending centrality’, in P.J. Carrington, J. Scott and S. Wasserman (eds), *Models and Methods in Social Network Analysis*, Cambridge, UK: Cambridge University Press, pp. 57–76.

Farmer, J.D., M. Shubik and E. Smith (2005), *Economics: The Next Physical Science?*, Working Paper # 05-06-27, Santa Fe Institute, Santa Fe, New Mexico http://www.santafe.edu/research/publications/wpabstract/200506027.

Gkantsidis, C., M. Mihail and E. Zegura (2003), ‘Spectral analysis of Internet topologies’, in Proceedings of Infocom 2003, http://citeseer.ist.psu.edu//gkantsidis03spectral.html.

Gladwell, M. (2000), *The Tipping Point: How Little Things Can Make a Big Difference*, Boston, MA, USA: Little, Brown.

Goodchild, M.F. (1992), 'Geographic information science', *International Journal of Geographical Information Systems*, **6**, 31–45.

Gorman, S.P. and R. Kulkarni (2004), 'Spatial small worlds: new geographic patterns for an information economy', *Environment and Planning B: Planning and Design*, **31** (2), 273–96.

Granovetter, M.S. (1973), 'The strength of weak ties', *American Journal of Sociology*, **78**, 1360–80.

Gross, J.L. and J. Yellen (eds) (2004), *Handbook of Graph Theory*, Boca Raton, FL, USA: CRC Press.

Haggett, P. and R.J. Chorley (1969), *Network Analysis in Geography*, London, UK: Arnold.

Hillier, B. and J. Hanson (1984), *The Social Logic of Space*, Cambridge, UK: Cambridge University Press.

Jacobs, J. (1961), *The Death and Life of Great American Cities: The Failure of Town Planning*, New York, NY, USA: Random House.

Jones, A. (1971), 'An order neighbor approach to random disturbances on regular point lattices', *Geographical Analysis*, **3**, 361–78.

Jumelius, E. and T. Petersen (2005), 'Importance and exposure in road network vulnerability analysis: a case study of northern Sweden', paper presented at the 8th NECTAR Conference, Las Palmas, Gran Canaria, June 1–4, 2005, CD-ROM.

Kansky, K.J. (1963), *Structure of Transport Networks: Relationships Between Network Geometry and Regional Characteristics*, University of Chicago, Department of Geography, research paper #84.

Kitchenham, B. (1990), 'Software development metric and models', in P. Rook (ed.), *Software Reliability Handbook*, New York, NY, USA: Elsevier Applied Science, pp. 441–86.

Kleinfeld, J. (2002), 'Six degrees of separation: urban myth?', *Psychology Today*, **35** (2), 74.

Lakhina, A., J.W. Byers, M. Crovella and I. Matta (2002), 'On the geographic locations of Internet resources', in Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement, Session 8 IGP and Topology. Also: http://www.cs.bu.edu/techreports/pdf/2002-015-internet-geography.pdf.

Lakhina, A., J.W. Byers, M. Crovella and P. Xie (2003), *Sampling Biases in IP Topology Measurements*, IEEE Infocom 2003. Also: Boston University Computer Science, Tech. Rep. BUCS-TR-2002-021, July 2002, http://www.ieee-infocom.org/2003/papers/09_01.PDF/lakhina03sampling.pdf.

Lea, A.C. and N.M. Waters (1996), 'The role of transportation in the Canadian economy', in J.N. Britton (ed.), *Canada and the Global Economy: The Geography of Structural and Technological Change*, Montreal and Kingston, Canada: McGill-Queen's University Press, pp. 299–315.

Lee, D.-H. (ed.) (2004), *Urban and Regional Transportation Modeling: Essays in Honor of David Boyce*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Lee, K. and H.-Y. Lee (1998), 'A new algorithm for graph theoretic nodal accessibility measurement', *Geographical Analysis*, **30** (1), 1–14.

Lee, P.J. and D. Gill (1999), ‘Comparison of discovery process methods for estimating undiscovered resources’, *Bulletin of Canadian Petroleum Geology*, **47** (1), 19–30.

Lee, P.J. and P.C.C. Wang (1986), ‘Evaluation of petroleum resources from pool size distributions’, in D.D. Rice (ed.), *Oil and Gas Assessments, Methods and Applications*, *AAPG Studies in Geology*, **21**, 33–42.

Longley, P.A., M.F. Goodchild, D.J. Maguire and D.W. Rhind (2001), *Geographic Information: From Systems to Science*, New York, NY, USA: Wiley.

Milgram, S. (1967), ‘The small world problem’, *Psychology Today*, **1**, 61–7.

Moore, G.A. (1991), *Crossing the Chasm: Marketing and Selling High Tech Products to Mainstream Customers*, New York, NY, USA: Harper Business.

Nagurney, A. (ed.) (2003), *Innovations in Financial and Economic Networks, Cheltenham, UK and Northampton*, MA, USA: Edward Elgar.

Nagurney, A. and J. Dong (2002), *Supernetworks: Decision-Making for the Information Age*, Cheltenham, UK and Northampton, MA, USA: Edward Elgar.

Nijkamp, P. and A. Reggiani (1989), ‘Spatial interaction and input-output models: a dynamic stochastic multiobjective framework’, in R.E. Miller, K. Polenske and A.Z. Rose (eds), *Frontiers of Input–Output Analysis*, Oxford, UK: Oxford University Press, pp. 193–205.

de Nooy, W., A. Mrvar and V. Batagelj (2005), *Exploratory Social Network Analysis with Pajek*, Cambridge, UK: Cambridge University Press.

Nystuen, J. and M. Dacey (1961), ‘A graph theory interpretation of nodal regions’, *Papers of the Regional Science Association*, **7**, 29–42.

O’Hagan, S.B. and M.B. Green (2004), ‘Corporate knowledge transfer via interlocking directorates: a network analysis approach’, *Geoforum*, **35** (1), 127–39.

Oldfield, S. (2004), ‘Urban networks, community organising and race: an analysis of racial integration in a desegregated South African neighbourhood’, *Geoforum*, **35** (2), 189–201.

Palla, G., I. Derenyi, I. Farkas and T. Vicsek (2005), ‘Uncovering the overlapping community structure of complex networks in nature and society’, *Nature*, **435**, 814–18.

Porta, S., P. Crucitti and V. Latora (2005), *The Network Analysis of Streets: A Primal Approach*, http://www.unifr.ch/econophysics, June.

Reed, W.R. (1970), ‘Indirect connectivity and hierarchies of urban dominance’, *Annals of the Association of American Geographers*, **60**, 770–85.

Rogers, E.M. (2003), Diffusion of Innovations, 5th edn, New York, NY, USA: Free Press.

Rosenthal, H. (1960), ‘Acquaintances and contacts of Franklin Roosevelt’, Master’s Thesis, MIT, Cambridge, MA, USA.

Roush, W. (2004), ‘Search beyond Google’, *Technology Review*, **107** (2), 34–45.

Schintler, L.A., S.P. Gorman, A. Reggiani, R. Patuelli, A. Gillespie, P. Nijkamp and J. Rutherford (2005b), ‘Complex network phenomena in telecommunication systems’, *Networks and Spatial Economics*, **5** (4), 351–70.

Schintler, L.A., S.P. Gorman, A. Reggiani, R. Patuelli and P. Nijkamp (2005a), ‘Scale-free phenomena in communications networks: a cross-Atlantic comparison’, in A. Reggiani and L.A. Schintler (eds), *Methods and Models in Transport and Telecommunications. Cross Atlantic Perspectives*, Berlin and Heidelberg, Germany and New York, NY, USA: Springer, pp. 201–20.

Seel, K. (2004), 'Boom and bust cycles in power plant construction: A simulation study of the temporal and geographical aspects of the Alberta competitive electrical industry', PhD Dissertation, Department of Geography, University of Calgary, Calgary, Canada.

Song, C., S. Havlin and H.A. Makse (2005), 'Self-similarity of complex networks', *Nature*, **433**, 392–5.

Taaffe, E.J. and H.L. Gauthier (1973), *Geography of Transportation*, Englewood Cliffs, NJ, USA: Prentice-Hall.

Taaffe, E.J., H.L. Gauthier and M.E. O'Kelly (1996), *Geography of Transportation*, 2nd edn, Upper Saddle River, NJ, USA: Prentice-Hall.

Tangmunarunkit, H., R. Govindan, S. Jamin, S. Sheker and W. Willinger (2001), *Network Topologies, Power Laws and Hierarchy*, Technical Report 01-746, Computer Science Department, University of Southern California, June, http://citeseer.ist.psu.edu/tangmunarunkit01network.html.

Thrift, N. (2002), 'The future of geography', *Geoforum*, **33** (3), 291–8.

Travers, J. and S. Milgram (1969), 'An experimental study of the small world problem', *Sociometry*, **32** (4), 425–43.

Waldinger, R. and D. Fitzgerald (2004), 'Transnationalism in question', *American Journal of Sociology*, **109** (5), 1177–95.

Warntz, W. (1973), 'New geography as general spatial systems theory – old social physics writ large?', in R.J. Chorley (ed.), *Directions in Geography*, London, UK: Methuen.

Wasserman, S. and K. Faust (1994), *Social Network Analysis: Methods and Applications*, Cambridge, UK: Cambridge University Press.

Waters, N.M. (1973), 'The growth of the southern Ontario railway system – a network analysis', MA thesis, Department of Geography, University of Western Ontario.

Waters, N.M. (1993a), 'GIS: paradigms lost', *GIS World*, **6** (8), 64.

Waters, N.M. (1993b), 'GIS and the 5,000 year bug: solving software reliability problems', *GIS World*, **6** (2), 64.

Waters, N. (2001), 'GIS, the Internet and six degrees of separation', *GEOWorld*, **14** (1), 28–9, http://www.geoplace.com/gw/2001/0101/0101edge.asp.

Waters, N.M. (2003), 'What can GIS do to save geography?', *GEOWorld*, **16** (12), 22–3, http://www.geoplace.com/gw/2003/0312/0312ends.asp.

Watts, D.J. (1999), *Small Worlds: The Dynamics of Networks between Order and Randomness*, Princeton, NJ, USA: Princeton University Press. Also http://smallworld.columbia.edu/.

Watts, D.J. (2003), *Six Degrees: The Science of a Connected Age*, New York, NY, USA: Norton.

Watts, D.J. and S.H. Strogatz (1998), 'Collective dynamics of 'small world' networks', *Nature*, **393**, 440–42.

Wheeler, D.C. and M.E. O'Kelly (1999), 'Network topology and city accessibility of the commercial Internet', *Professional Geographer*, **51** (3), 327–39.

Wilson, A.G. (1970), *Entropy in Urban and Regional Modelling*, London, UK: Pion.

Wilson, E.O. (1998), *Consilience: The Unity of Knowledge*, New York, NY, USA: Knopf.