ELSEVIER

Social Networks 22 (2000) 187–220

SOCIAL NETWORKS

www.elsevier.com/locate/socnet

Dynamical systems to define centrality in social networks

R. Poulin [a,b], M.-C. Boily [a,b,*], B.R. Mâsse [a,b]

[a] *CHA, Pavillon St. Sacrement, Groupe de Recherche en Épidémiologie, 1050 Chemin Ste-Foy, Québec, Québec, Canada G1S 4L8*

[b] *Département de Médecine Sociale et Préventive, Université Laval, Québec, Canada*

Abstract

In this paper, new measures of centrality that summarize the contact structure of social networks are proposed. The new measures use a cumulative nomination scheme based on the preliminary assumption that more central individuals will be nominated more often. Some of these measures are defined to characterize networks of different sizes and, by extension, networks made of many components. These new measures are applied to a network of 40 homosexuals with AIDS [Auerbach, D., Darrow, W., Jaffe, H., Curran, J., 1984. Cluster of cases of the acquired immune deficiency syndrome: patients linked by sexual contact. Am. J. Med. 76 (1984) 487–492; Klovdahl, A.S., 1985. Social networks and the spread of infectious diseases: the AIDS example. Soc. Sci. Med. 21 (1985) 1203–1216.], an illustrative multi-component network and a simulated network. They are compared to classical measures based on geodesics (closeness, eccentricity), to information-based centrality measures introduced by Stephenson and Zelen [Stephenson, K., Zelen, M., 1989. Rethinking centrality: methods and examples. Soc. Networks 11 (1989) 1–37.] and Altmann [Altmann, M., 1993. Reinterpreting network measures for models of disease transmission. Soc. Networks 15 (1993) 1–17.], and to the centrality measure of Bonacich [Bonacich, P., 1972. Factoring and weighting approaches to status scores and clique identification. J. Math. Sociol. 2 (1972) 113–120.]. The most basic of our measures is shown to be related to the Bonacich index of centrality for connected networks. The scaling law of the different centrality measures is examined by measuring simulated networks of various sizes. Measures based on the distribution of the components' size obey a simple proportional scaling law while those based on geodesics do not. Our new measures prove interesting because they consider all the possible paths, do not require intensive computer calculations, and can be used to compare networks of different sizes because they are independent of the size of the networks. © 2000 Elsevier Science B.V. All rights reserved.

* Corresponding author. Département de Médecine Sociale et Préventive, Université Laval, Québec, Canada G1K 7P4. Tel.: +1-418-656-3333.

E-mail address: mcboily@gre.ulaval.ca (M.-C. Boily).

0378-8733/00/$ - see front matter
PII: S0378-8733(00)00020-4

1. Introduction

In the past 6 years, the study of social networks has gained in popularity among sexually transmitted diseases (STD) epidemiologists. This is the result of two phenomena: sociologists moving to the field of STD/human immunodeficiency virus (HIV) epidemiology with their own tools and STD/HIV epidemiologists' increasing interest for the structural organization of relationships, following results on the importance of the mixing patterns on the course of infections (Anderson, 1988, 1991; Anderson et al., 1989a,b, 1990; 1991; Boily and Anderson, 1991). There is now a growing consensus among STD/HIV epidemiologists and sociologists that the more intricate nature of sexual and needle exchange partnerships formation and the finer network structure may also play a determinant role on disease transmission (Altmann et al., 1994; Rothenberg et al., 1995, 1998; Iannelli et al., 1997).

In a first generation of models of STD and HIV transmission, heterogeneity in sexual activity and the mixing pattern between different risk groups (age, sexual activity classes, etc.) have been shown to play an important role on the establishment, the time course of the epidemic and the equilibrium prevalence of the infection (Anderson et al., 1986; Hethcote and Yorke, 1984). A second generation of models, including the dynamics of partnerships formation (Dietz, 1988), and observational studies (Klovdahl et al., 1994) have shown that other structural measures of networks are also relevant to the dynamics of transmission (Kretzschmar and Wiessing, 1998; Friedman et al., 1997) in population and to explain individual risks of infection.

Some studies have considered global measures of networks to summarize the density of a network, e.g., the information on the connectivity. Other measures attribute a quantity to each individual to reflect his position within the network. The latter type of measure leads to the concept of centrality where individuals are considered to be central if they occupy pivotal positions with respect to the global network structure. The quantification of centrality of an individual has been performed by Freeman (1979) (see also Wasserman and Faust (1994) for a comprehensive review on centrality) based on geodesics (i.e., shortest paths) between individuals. However, since centrality measures based on geodesics do not consider all the possible paths between individuals, it can be argued that this type of measure does not capture the finest structure of a network.

An index of centrality, which has been shown to be equivalent to the conductance in an electrical network (Altmann, 1993), has been introduced by Stephenson and Zelen (1989). Their measure, based on the quantity of transmissible information between any pair of individuals in a connected network, considers all possible paths and captures the finest structure of the network. The calculation of their measure is based on the inverse of an $N \times N$ matrix where N is the size of the connected network.

Another index of centrality, which also considers all possible paths, has been proposed by Bonacich (1972). The idea is to attribute a popularity score to each individual within a connected network. To obtain these popularity scores, one of the methods is to solve a homogeneous set of N linear equations. The problem can be formulated as an eigensystem from which the eigenvector corresponding to the largest eigenvalue is proportional to the desired popularity scores. A second method uses a mapping which is iterated to converge to this eigenvector.

The measure of Stephenson–Zelen (S–Z), or the measure of Bonacich, can be used to characterize small connected networks. For large networks, however, the inversion of the matrix for S–Z, or the solution of the eigensystem for Bonacich, rapidly becomes untractable as N increases, even for powerful computers. For Bonacich's mapping, the convergence is often very slow and not always guaranteed. Moreover, for networks composed of many components (or networks of different sizes), comparisons of individuals between components, based on Bonacich's, Stephensen and Zelens' or Freeman's centrality measures, become problematic.

In this paper, a set of new centrality measures, also based on the iteration of a mapping, is proposed. For connected networks, the most basic of our set of measures is shown to be related to the Bonacich index of centrality. The new centrality measures consider all the possible paths between pairs of individuals, capture the finest structure of the network, and are suitable to characterize large networks of variable size or made of many components because they are easy to implement and are not computer-intensive as the mapping on which they are based converges rapidly. The validity of these new measures will be evaluated within and between components of different sizes and compared to other centrality measures. In this paper, we consider sexual networks, but the generalization of the results to any undirected relational network is straightforward.

The rest of the paper is structured as follows. In Section 2, a short review of the required basic graph terminology is presented. In Section 3, an overview of some of the principal structural measures, such as the distribution of the size of the components (in multi-components networks), measures based on geodesics (Freeman's measures of eccentricity and closeness), measures based on information (S–Z's and Altmann's versions), and the Bonacich measure is given. Along with their definitions, the different measures are applied to two different networks in order to illustrate the meaning of each measure. The first network is made of many small components while the second is made of a single component and represents a contact network of 40 homosexual men with AIDS (Auerbach et al., 1984; Klovdahl, 1985). In Section 4, the centrality indices based on mappings are presented. The mapping proposed by Bonacich is first introduced and slightly modified to obtain a better numerical stability. The behavior of his mapping is studied and its pathological properties identified. A new mapping, based on a generic cumulative nomination scheme, is also proposed. Computationally, our mapping is shown to be superior to Bonacich's even though both mappings produce, to a normalization constant, similar results for connected networks. Centrality indices suitable to characterize points from different components are then introduced. Again, along with their definitions, the different measures will be applied to the two illustrative networks and compared to more classical centrality measures. Finally, in Section 5, the various structural measures (the classical ones and those proposed in this paper) are validated by examining and comparing different centrality indices for the case of a connected sexual network of 40 homosexuals with AIDS (Auerbach et al., 1984; Klovdahl, 1985), an illustrative multi-component network and a simulated network, and by examining their scaling law (behaviour of the measures as N varies) for simulated multi-component networks.

2. Some basic graph terminology

In this section, only the graph terminology required for the understanding of this paper is presented. A more exhaustive review of terminology and theory can be found, in Harary (1969), Freeman (1979) and Wasserman and Faust (1994).

Consider the network (or graph) shown in Fig. 1. It is made of 50 points (or nodes or vertices) and 51 lines (or links or edges or arcs). For the sake of simplicity, we shall limit ourselves to undirected and unweighted networks. In this case, if a point *i* is linked to *j*, then *j* is necessarily linked to *i*. Moreover, all the links have the same weight. This generalization includes a wide range of networks such as sexual networks. The degree of a point is the number of points to which it is adjacent, i.e., directly linked. For example, in Fig. 1, point 4 has degree 1 and point 5 has degree 2. Point 4 is adjacent to 5 but not to 6. However, point 6 can be reached by 4 following the path 4–5–6. A network is said to be connected if each point can be reached from any other point. The network presented in Fig. 1 is not connected and is made of 13 components, noted A–M, with sizes varying from 1 to 5. Different names have been given by graph theorists to represent the configuration of particular networks. Let us consider the network of Fig. 1 as a collection of smaller networks (components). The name *line graph* would then apply to components B, C, E and I; *circle graph* to components D, F and J; *star graph* to components C, G and K; and finally *complete graph* to components B, D, H and M, because each point is adjacent to each other.

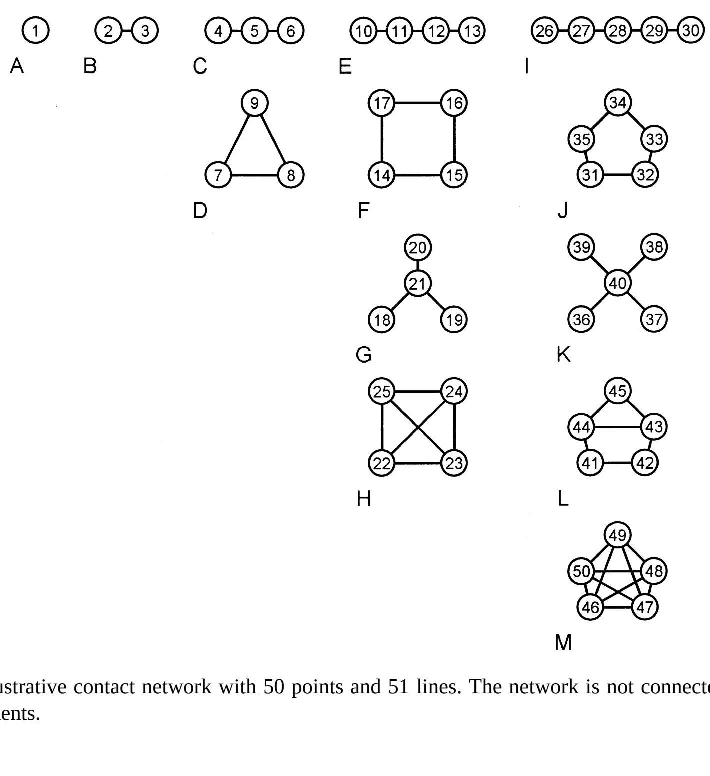

Fig. 1. An illustrative contact network with 50 points and 51 lines. The network is not connected but is made of 13 components.

Mathematically, it is usually convenient to describe a network using an adjacency matrix (sociomatrix in sociology) **A** whose entries are given by:

$$a_{ij} = \begin{cases} 1 & \text{if } i \text{ is adjacent to } j, \\ 0 & \text{if not.} \end{cases}$$

Since the study is limited to undirected contact networks, $a_{ii} = 0$ and **A** is symmetric, i.e., $a_{ij} = a_{ji}$.

3. Structural measures and centrality

In this section, we present a selection of classical measures typically used to characterize the structure of networks. An ideal set of measures should be easy to interpret, discriminant and independent of each other.

3.1. Components' size distribution

A simple way to characterize the structure of a network is to consider the distribution of the size of its components. For example, in Fig. 1, the network is made of one component of size 1, one of size 2, two of size 3, four of size 4 and five of size 5. Different quantities can be used to summarize the information contained in the complete distribution. In this paper, we consider the following measures:

N the total size of the network (number of points)
NC the number of components
CMax the size of the biggest component
CMean the mean component size (N/NC)
CDev the standard deviation of the components' size distribution

For the network of Fig. 1, we obtain : $N = 50$, NC = 13, CMax = 5, CMean = 3.84 and CDev = 1.28.

Another approach is to consider the distribution from a node point of view. Let CSize(i) be the size of the component within which point i is located, relative to the total size of the network (N). In Fig. 1, the network is formed of one point for which CSize equals 0.02 (1/50), two points with CSize = 0.04, six points with CSize = 0.06, 16 points with CSize = 0.08 and 25 points with CSize = 0.10. Although a large variety of statistics can be used to describe the distribution, the following quantities will be considered:

MCSize the mean of the distribution CSize(i)
SDCSize the standard deviation of CSize(i)
ADCSize the absolute deviation of CSize(i)
MaxCSize the maximum of CSize(i)

For the network of Fig. 1, we obtain : MCSize = 0.0848, SDCSize = 0.0192, ADCSize = 0.0152 and MaxCSize = 0.1000.

3.2. Centrality and Freeman indices

The measures described above are based on network components treating each component as a global entity (each component = 1 observation). However, within a component, individuals are not necessarily all structurally equivalent. This observation led to the introduction of the popular concept of centrality (Freeman, 1979; Wasserman and Faust, 1994), which attempts to describe the structure of a connected network (formed of one component only). In simple words, the concept of centrality refers to the importance (or contribution) of a point (individual) to the global structure (or configuration) of the network. Loosely speaking, a point i is said to be central if it has some potential impact on the other points of the network. A point i in a network has a potential impact only on the points that are reachable, i.e., the points that are within the same component. A simple index of centrality for an individual i could then be:

$$C_{\mathrm{CSize}}(i) \equiv \mathrm{CSize}(i),$$

the relative size of the component within which i is located. Table 1 lists C_{CSize} corresponding to each point of the contact network of Fig. 1. However, to discriminate points within a component or within a network made of only one component, the simplest index of centrality is the degree of the point, i.e.,

$$C_{\mathrm{Deg}}(i) \equiv \mathrm{Degree}(i).$$

Consider the contact network illustrated in Fig. 2. It represents a sexual network of 40 homosexual men with AIDS (Auerbach et al., 1984; Klovdahl, 1985). Individual 16 gets the highest degree centrality score (8). It is intuitively clear from Fig. 2 that this individual could have an important impact on HIV transmission, for example. Table 2 lists individuals of Fig. 2 sorted in descending order according to their degree centrality index C_{Deg} (given in parentheses). Individual 16 is top rated with a score of 8, then 5 and 26 each gets a score of 5. Are these two individuals structurally equivalent? Should they be considered as important with respect to HIV transmission through other individuals in the network? Certainly not. It is rather intuitive to establish that individual 26 is more important while individual 5 is rather peripheral. The degree of a point is a local measure of the network around that point and does not thoroughly reflect its complete structure.

One way, suggested by Freeman (1979), to discriminate between points having the same degree is to consider a more global network approach. Let us consider the geodesic, i.e., the shortest path between two points in a connected network. For example, in component L of the network of Fig. 1, the shortest path between points 41 and 45 is the path 41–44–45 of length 2. Another path between 41 and 45 could be 41–42–43–45, a path of length 3. If we define the distance (geodesic) $d(i,j)$ as the length of the shortest path between points i and j in a connected network (Freeman, 1979), then in our example, $d(41,45) = 2$.

Based on the distance $d(i,j)$, a measure of centrality, called the eccentricity of point i, is expressed by (Wasserman and Faust, 1994):

$$C_{\mathrm{Ey}}(i) = \max_{j}\{d(i,j)\}/(N-1), \tag{1}$$

Table 1
Centrality measures for points in the contact network of Fig. 1

Point	Component	C_{CSize}	C_{Deg}	C_{Ey}	C_{Cs}	C_{Inf}	C'_{Inf}	C_{Bon}	C_{λ}	C_{CN}	C_{CNGr}	C'_{CN}	C''_{CN}
1	A	0.02	0	1.000	0.000	0.000	0.000	1.000	0.00	1.000	1.00	1.000	0.020
2	B	0.04	1	1.000	1.000	2.000	0.500	0.707	1.00	1.000	2.00	2.000	0.080
3	B	0.04	1	1.000	1.000	2.000	0.500	0.707	1.00	1.000	2.00	2.000	0.080
4	C	0.06	1	1.000	0.667	1.000	0.500	0.500	1.41	0.879	2.41	2.121	0.127
5	C	0.06	2	0.500	1.000	1.500	0.667	0.707	1.41	1.243	2.41	3.000	0.180
6	C	0.06	1	1.000	0.667	1.000	0.500	0.500	1.41	0.879	2.41	2.121	0.127
7	D	0.06	2	0.500	1.000	2.250	1.000	0.577	2.00	1.000	3.00	3.000	0.180
8	D	0.06	2	0.500	1.000	2.250	1.000	0.577	2.00	1.000	3.00	3.000	0.180
9	D	0.06	2	0.500	1.000	2.250	1.000	0.577	2.00	1.000	3.00	3.000	0.180
10	E	0.08	1	1.000	0.500	0.667	0.458	0.372	1.62	0.764	2.62	2.000	0.160
11	E	0.08	2	0.667	0.750	1.000	0.625	0.601	1.62	1.236	2.62	3.236	0.259
12	E	0.08	2	0.667	0.750	1.000	0.625	0.601	1.62	1.236	2.62	3.236	0.259
13	E	0.08	1	1.000	0.500	0.667	0.458	0.372	1.62	0.764	2.62	2.000	0.160
14	F	0.08	2	0.667	0.750	1.600	0.917	0.500	2.00	1.000	3.00	3.000	0.240
15	F	0.08	2	0.667	0.750	1.600	0.917	0.500	2.00	1.000	3.00	3.000	0.240
16	F	0.08	2	0.667	0.750	1.600	0.917	0.500	2.00	1.000	3.00	3.000	0.240
17	F	0.08	2	0.667	0.750	1.600	0.917	0.500	2.00	1.000	3.00	3.000	0.240
18	G	0.08	1	0.667	0.600	0.800	0.500	0.408	1.73	0.845	2.73	2.309	0.185
19	G	0.08	1	0.667	0.600	0.800	0.500	0.408	1.73	0.845	2.73	2.309	0.185
20	G	0.08	1	0.667	0.600	0.800	0.500	0.408	1.73	0.845	2.73	2.309	0.185
21	G	0.08	3	0.333	1.000	1.333	0.750	0.707	1.73	1.464	2.73	4.000	0.320
22	H	0.08	3	0.333	1.000	2.667	1.500	0.500	3.00	1.000	4.00	4.000	0.320
23	H	0.08	3	0.333	1.000	2.667	1.500	0.500	3.00	1.000	4.00	4.000	0.320
24	H	0.08	3	0.333	1.000	2.667	1.500	0.500	3.00	1.000	4.00	4.000	0.320
25	H	0.08	3	0.333	1.000	2.667	1.500	0.500	3.00	1.000	4.00	4.000	0.320
26	I	0.10	1	1.000	0.400	0.500	0.417	0.289	1.73	0.670	2.73	1.830	0.183
27	I	0.10	2	0.750	0.571	0.714	0.567	0.500	1.73	1.160	2.73	3.170	0.317
28	I	0.10	2	0.500	0.667	0.833	0.600	0.577	1.73	1.340	2.73	3.660	0.366
29	I	0.10	2	0.750	0.571	0.714	0.567	0.500	1.73	1.160	2.73	3.170	0.317
30	I	0.10	1	1.000	0.400	0.500	0.417	0.289	1.73	0.670	2.73	1.830	0.183
31	J	0.10	2	0.500	0.667	1.250	0.833	0.447	2.00	1.000	3.00	3.000	0.300
32	J	0.10	2	0.500	0.667	1.250	0.833	0.447	2.00	1.000	3.00	3.000	0.300
33	J	0.10	2	0.500	0.667	1.250	0.833	0.447	2.00	1.000	3.00	3.000	0.300
34	J	0.10	2	0.500	0.667	1.250	0.833	0.447	2.00	1.000	3.00	3.000	0.300
35	J	0.10	2	0.500	0.667	1.250	0.833	0.447	2.00	1.000	3.00	3.000	0.300
36	K	0.10	1	0.500	0.571	0.714	0.500	0.354	2.00	0.833	3.00	2.500	0.250
37	K	0.10	1	0.500	0.571	0.714	0.500	0.354	2.00	0.833	3.00	2.500	0.250
38	K	0.10	1	0.500	0.571	0.714	0.500	0.354	2.00	0.833	3.00	2.500	0.250
39	K	0.10	1	0.500	0.571	0.714	0.500	0.354	2.00	0.833	3.00	2.500	0.250
40	K	0.10	4	0.250	1.000	1.250	0.800	0.707	2.00	1.667	3.00	5.000	0.500
41	L	0.10	2	0.500	0.667	1.410	0.939	0.358	2.48	0.812	3.48	2.827	0.283
42	L	0.10	2	0.500	0.667	1.410	0.939	0.358	2.48	0.812	3.48	2.827	0.283
43	L	0.10	3	0.500	0.800	1.774	1.176	0.530	2.48	1.203	3.48	4.188	0.419
44	L	0.10	3	0.500	0.800	1.774	1.176	0.530	2.48	1.203	3.48	4.188	0.419
45	L	0.10	2	0.500	0.667	1.375	0.967	0.427	2.48	0.970	3.48	3.376	0.338
46	M	0.10	4	0.250	1.000	3.125	2.000	0.447	4.00	1.000	5.00	5.000	0.500
47	M	0.10	4	0.250	1.000	3.125	2.000	0.447	4.00	1.000	5.00	5.000	0.500
48	M	0.10	4	0.250	1.000	3.125	2.000	0.447	4.00	1.000	5.00	5.000	0.500
49	M	0.10	4	0.250	1.000	3.125	2.000	0.447	4.00	1.000	5.00	5.000	0.500
50	M	0.10	4	0.250	1.000	3.125	2.000	0.447	4.00	1.000	5.00	5.000	0.500

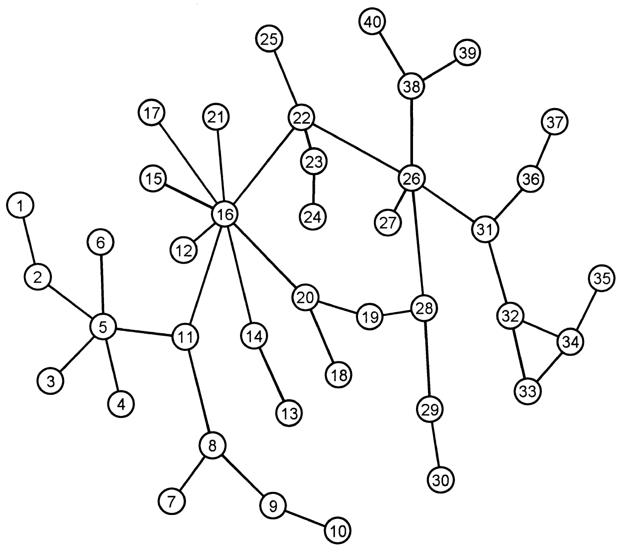

Fig. 2. Sexual network of 40 homosexual men with AIDS (from Fig. 2 in Stephenson and Zelen, 1989; Fig. 1 in Klovdahl, 1985; Fig. 1 in Auerbach et al., 1984).

and represents the maximal distance between the point i and any other point j in the network. This quantity is standardized such that the highest possible value for a connected network of size N is 1.

Back to our AIDS example, points 5 and 26 can now be discriminated (see Table 2) since $C_{\mathrm{Ey}}(5) = 0.205$ (rank 19) and $C_{\mathrm{Ey}}(26) = 0.154$ (rank 6). With respect to eccentricity, point 26 is considered more central than point 5 since it gets a smaller eccentricity score. Moreover, being located in the middle of the contact network, individual 22 is now considered to be the most central. However, if we consider component L of the network of Fig. 1, again, the global index of centrality given by the eccentricity does not discriminate points 41,…,45 with values of C_{Ey} all equal to 0.500.

Perhaps the most popular in the literature, the closeness (Rothenberg et al., 1995; Kretzschmar et al., 1995; Wasserman and Faust, 1994; Freeman, 1979), is another global measure of centrality based on the distance between points in a connected network. The closeness of point i is given by the following expression:

$$C_{\mathrm{Cs}}(i) = (N-1) \Big/ \sum_j d(i,j). \qquad (2)$$

The measure is standardized such that, for a connected network of size N, the maximal possible value of $C_{\mathrm{Cs}}(i)$ is one. More central points should score higher on the closeness scale. For example, in component L (Fig. 1), points 43 and 44 are more central since $C_{\mathrm{CS}}(43, 44) = 0.800$ and $C_{\mathrm{Cs}}(41, 42, 45) = 0.667$ (cf. Table 1). For the AIDS network (cf. Fig. 2 and Table 2), individual 26 is now ranked 3rd with a closeness centrality index of 0.322 and individual 16 comes back to the top position with a score of 0.351.

Table 2
Centrality measures for the sexual network presented in Fig. 2
For each measure, individuals are ranked in descending order of centrality from the most central (rank 1) to the less (rank 40). The corresponding centrality indices are given in parentheses. Centrality indices not listed in the table but presented in the paper are $\mathrm{CSize}(i) = 1$, $C_\lambda(i) = 3.117$, $C_{\mathrm{CNGr}}(i) = 4.117$ and $C''_{\mathrm{CN}}(i) = C'_{\mathrm{CN}}(i)$ for all individuals i since the network is made of only one component.

Rank	C_{Deg}	C_{EY}	C_{CS}	C_{Inf}	C'_{Inf}	C_{Bon}	C_{CN}	C'_{CN}
1	16 (8)	22 (0.128)	16 (0.351)	16 (0.417)	16 (0.516)	16 (0.587)	16 (5.252)	16 (21.62)
2	5 (5)	16 (0.154)	22 (0.345)	22 (0.392)	26 (0.472)	22 (0.339)	22 (3.039)	22 (12.51)
3	26 (5)	19 (0.154)	26 (0.322)	26 (0.388)	22 (0.466)	11 (0.273)	11 (2.441)	11 (10.05)
4	22 (4)	23 (0.154)	11 (0.302)	20 (0.357)	20 (0.425)	20 (0.254)	20 (2.275)	20 (9.366)
5	8 (3)	25 (0.154)	20 (0.281)	11 (0.351)	28 (0.418)	26 (0.241)	26 (2.161)	26 (8.900)
6	11 (3)	26 (0.154)	14 (0.265)	28 (0.348)	11 (0.406)	14 (0.210)	14 (1.878)	14 (7.732)
7	20 (3)	11 (0.179)	19 (0.265)	19 (0.336)	19 (0.394)	12 (0.188)	12 (1.685)	12 (6.936)
8	28 (3)	12 (0.179)	31 (0.265)	31 (0.310)	31 (0.372)	15 (0.188)	15 (1.685)	15 (6.936)
9	31 (3)	14 (0.179)	12 (0.262)	14 (0.303)	5 (0.360)	17 (0.188)	17 (1.685)	17 (6.936)
10	32 (3)	15 (0.179)	15 (0.262)	12 (0.299)	38 (0.346)	21 (0.188)	21 (1.685)	21 (6.936)
11	34 (3)	17 (0.179)	17 (0.262)	15 (0.299)	14 (0.345)	5 (0.152)	5 (1.358)	5 (5.592)
12	38 (3)	20 (0.179)	21 (0.262)	17 (0.299)	32 (0.332)	28 (0.132)	28 (1.185)	28 (4.881)
13	2 (2)	21 (0.179)	23 (0.262)	21 (0.299)	12 (0.329)	19 (0.124)	19 (1.110)	19 (4.570)
14	9 (2)	24 (0.179)	28 (0.262)	38 (0.292)	15 (0.329)	23 (0.121)	23 (1.087)	23 (4.475)
15	14 (2)	27 (0.179)	25 (0.258)	23 (0.290)	17 (0.329)	8 (0.112)	8 (1.001)	8 (4.121)
16	19 (2)	28 (0.179)	38 (0.252)	25 (0.285)	21 (0.329)	25 (0.109)	25 (0.975)	25 (4.014)
17	23 (2)	31 (0.179)	5 (0.248)	27 (0.283)	8 (0.327)	31 (0.106)	31 (0.947)	31 (3.900)
18	29 (2)	38 (0.179)	27 (0.245)	5 (0.282)	23 (0.327)	38 (0.098)	38 (0.873)	38 (3.595)
19	33 (2)	5 (0.205)	8 (0.242)	8 (0.274)	27 (0.312)	18 (0.082)	18 (0.730)	18 (3.004)
20	36 (2)	8 (0.205)	18 (0.220)	18 (0.266)	25 (0.311)	27 (0.077)	27 (0.693)	27 (2.855)
21	1 (1)	13 (0.205)	32 (0.218)	29 (0.265)	29 (0.305)	13 (0.067)	13 (0.602)	13 (2.480)
22	3 (1)	18 (0.205)	36 (0.213)	32 (0.248)	34 (0.298)	2 (0.054)	2 (0.486)	2 (2.000)
23	4 (1)	29 (0.205)	13 (0.211)	36 (0.242)	18 (0.292)	32 (0.051)	32 (0.453)	32 (1.864)
24	6 (1)	32 (0.205)	29 (0.211)	13 (0.236)	33 (0.288)	3 (0.049)	3 (0.436)	3 (1.794)
25	7 (1)	36 (0.205)	24 (0.209)	39 (0.228)	36 (0.284)	4 (0.049)	4 (0.436)	4 (1.794)
26	10 (1)	39 (0.205)	2 (0.202)	40 (0.228)	2 (0.272)	6 (0.049)	6 (0.436)	6 (1.794)
27	12 (1)	40 (0.205)	39 (0.202)	24 (0.227)	9 (0.260)	29 (0.047)	29 (0.424)	29 (1.745)
28	13 (1)	2 (0.231)	40 (0.202)	2 (0.225)	13 (0.257)	9 (0.040)	9 (0.358)	9 (1.474)
29	15 (1)	3 (0.231)	3 (0.200)	3 (0.222)	3 (0.256)	24 (0.039)	24 (0.349)	24 (1.435)
30	17 (1)	4 (0.231)	4 (0.200)	4 (0.222)	4 (0.256)	36 (0.038)	36 (0.339)	36 (1.395)
31	18 (1)	6 (0.231)	6 (0.200)	6 (0.222)	6 (0.256)	7 (0.036)	7 (0.321)	7 (1.322)
32	21 (1)	7 (0.231)	9 (0.198)	9 (0.220)	39 (0.254)	39 (0.031)	39 (0.280)	39 (1.153)
33	24 (1)	9 (0.231)	7 (0.196)	7 (0.218)	40 (0.254)	40 (0.031)	40 (0.280)	40 (1.153)
34	25 (1)	30 (0.231)	34 (0.182)	34 (0.217)	24 (0.247)	34 (0.027)	34 (0.242)	34 (0.995)
35	27 (1)	33 (0.231)	33 (0.181)	33 (0.216)	7 (0.243)	33 (0.025)	33 (0.223)	33 (0.917)
36	30 (1)	34 (0.231)	37 (0.176)	30 (0.212)	30 (0.233)	1 (0.017)	1 (0.156)	1 (0.641)
37	35 (1)	37 (0.231)	30 (0.175)	37 (0.197)	37 (0.221)	30 (0.015)	30 (0.136)	30 (0.560)
38	37 (1)	1 (0.256)	1 (0.169)	1 (0.185)	35 (0.214)	10 (0.013)	10 (0.115)	10 (0.473)
39	39 (1)	10 (0.256)	10 (0.166)	10 (0.182)	1 (0.212)	37 (0.012)	37 (0.109)	37 (0.447)
40	40 (1)	35 (0.256)	35 (0.155)	35 (0.180)	10 (0.205)	35 (0.009)	35 (0.077)	35 (0.319)

Note that the inverse of C_{Cs} is simply the mean distance between point i and any other point in a connected network (Rothenberg et al., 1995).

If the network is not connected, i.e., made of many components, the distance $d(i,j)$ between two points from different components is not defined. One could still obtain a centrality index by considering each component separately. However, as raised by Donninger (1986) and also discussed by Stephenson and Zelen (1989) and Altmann (1993), this approach raises the problem of graph comparability. How can the centrality indices of points in networks (components) of different sizes be compared? For example, it is clear that the quantities $\max_j \{d(i,j)\}$ and $\sum_j d(i,j)$ increase as the size of the network increases. However, what is less trivial is if the functional relationship between one of these quantities and the size of the network has any simple form (e.g., directly proportional). In other words, what is the scaling law between C_{Ey} or C_{Cs} (or any other structural measures) and N? This important question will be addressed in Section 5.2.

A closer look at component L of Fig. 1 shows that comparisons of measures based on geodesics cannot discriminate between points 41, 42 and 45, yet in different geometrical or structural positions. One solution to this problem is to consider all the possible paths between any pair of points in the connected network. For example, let us apply this idea and enumerate all the possible paths between the pair (41,44) and (45,44). We obtain three paths between points 41 and 44 : 41–44, 41–42–43–44 and 41–42–43–45–44 of length 1, 3 and 4, respectively, and three paths between the pair (45,44): 45–44, 45–43–44 and 45–43–42–41–44 of length 1, 2 and 4, respectively. From the structural point of view of node 44, when all possible paths are considered, points 41 and 45 are not equivalent. Path enumeration, however, is a very tedious task for large networks, even for powerful computers.

3.3. S–Z index of centrality

A structural measure that does not require path enumeration, but still considers all the possible paths, has been proposed by Stephenson and Zelen (1989). Their measure, based on the information that can be transmitted between any two points in a connected network, has been shown to be equivalent to the electrical conductance in an electrical network (Altmann, 1993).

Let **A** be an adjacency matrix describing the connected network, **D** a diagonal matrix of the degree of each point and **J** a matrix with all its elements equal to one. The index of centrality of S–Z is calculated by inverting the matrix **B** defined by:

$$\mathbf{B} = \mathbf{D} - \mathbf{A} + \mathbf{J},$$

in order to obtain the matrix:

$$\mathbf{C} = (c_{ij}) = \mathbf{B}^{-1}$$

from which the information matrix is given explicitly by:

$$\mathbf{I}_{ij} = \left(c_{ii} + c_{jj} - 2c_{ij}\right)^{-1}. \quad (3)$$

The values $\mathbf{I}_{ij}$ summarize the information contained in all possible paths between points i and j.

To define the centrality index associated with the point i, S–Z use the harmonic average:

$$C_{\mathrm{Inf}}(i) = \left[\frac{1}{N} \sum_{j} \frac{1}{\mathbf{I}_{ij}} \right]^{-1}, \tag{4}$$

without much justification for this choice, if not for computational purpose ($\mathbf{I}_{ii}$ is infinity, thus $1/\mathbf{I}_{ii} = 0$ and does not contribute to the harmonic average), over the more simple arithmetic average. As raised by Altmann (1993), the use of the harmonic average may be misleading because the contribution to the measure $C_{\mathrm{Inf}}(i)$ of pairs (i,j) between which a lot of information is transmitted will be negligible. Altmann thus proposed a different approach to obtain the information matrix where $\mathbf{I}_{ii}$ is finite and then define an information centrality index based on arithmetic average. We propose, however, to simply consider $\mathbf{I}_{ii}$ as undefined and use the information matrix (Eq. 3) to defined a centrality index based on the arithmetic average as:

$$C'_{\mathrm{Inf}}(i) = \frac{1}{N} \sum_{j \neq i} \mathbf{I}_{ij}. \tag{5}$$

Based on the S–Z information centrality index, C_{Inf}, points 41 and 42 can now be discriminated from point 45 (Fig. 1), the latter being considered less central (smaller score) since $C_{\mathrm{Inf}}(41) = C_{\mathrm{Inf}}(42) = 1.410$ and $C_{\mathrm{Inf}}(45) = 1.375$ (see Table 1). When the arithmetic average is used, the scores of the information centrality index are $C'_{\mathrm{Inf}}(41) = C'_{\mathrm{Inf}}(42) = 0.939$ and $C'_{\mathrm{Inf}}(45) = 0.967$. This new index also discriminates points 41–42 from point 45 except that point 45 is now considered more central. Intuitively, the latter result appears more adequate if we compare distances of all the possible paths. For example, the total distance of all the paths between points 41 and 44 is $1 + 3 + 4 = 8$ while that of pair (44,45) is $1 + 2 + 4 = 7$. From the point of view of 44, this suggests that point 45 should be more central than point 41 since it seems more easily accessible.

Stephenson and Zelen applied their information centrality index to the case of the AIDS network (Fig. 2). We have reproduced the same analysis using their measure (C_{Inf}) and Altmann version (C'_{Inf}). Results are presented in Table 2. Both measures rank individual 16 first, but individuals 22 and 26 are ranked differently. Index C_{Inf} ranks the first three individuals (16–22–26) the same way C_{Cs} does. More interesting is the fact that both C_{Inf} and C'_{Inf} discriminate individuals 14, 19, 31. This was not the case with the closeness for which $C_{\mathrm{Cs}}(14\text{–}19\text{–}31) = 0.265$.

Despite the good discriminant power of information-based measures (the harmonic or arithmetic version), their use with large networks may be computationally limited by the fact that they require to inverse a large matrix. Another shortcoming is that the effect of the network size on the scale of the measure is unknown. Based on results from percolation theory (Doyle and Snell, 1984), Altmann (1993) suggested that if the network is roughly two-dimensional, it may be better to rescale the sums in Eqs. 4 and 5 by $\sqrt{N}$ instead of N. In general, however, the networks of interest here (sexual or needle sharing) are not two-dimensional (not planar graph).

3.4. Bonacich index of centrality

Another index of centrality that considers all possible paths and does not require path enumeration has been proposed by Bonacich (1972). This index characterizes individuals in connected networks according to their level of popularity. A given individual is said to be popular if he is related to many individuals or to few individuals who are themselves popular. Mathematically, this can be written as the following set of homogeneous linear equations (Wasserman and Faust, 1994; Bonacich, 1972, 1987):

$$s_i = a_{i1} s_1 + \ldots + a_{iN} s_N, \tag{6}$$

where $s_i \geq 0$ denotes the popularity score of individual i, and a_{ij} is the entry of the adjacency matrix representing the network of contact (note that $a_{ii} = 0$).

The system of Eq. 6, which can be written in a matrix form as $S = \mathbf{A} S$ or $(\mathbf{A} - \mathbf{I})S = 0$, has a nonzero solution only when det $(\mathbf{A} - \mathbf{I}) = 0$ (which is not the case in general). However, by multiplying the left-hand side of Eq. 6 by some constant λ, we can always obtain a nonzero solution for:

$$\lambda s_i = a_{i1} s_1 + \ldots + a_{iN} s_N, \tag{7}$$

which, once written in matrix form as $\lambda S = \mathbf{A} S$, is nothing but the familiar problem of solving the eigensystem:

$$(\mathbf{A} - \lambda \mathbf{I}) S = 0. \tag{8}$$

For symmetrical matrices of contact, Bonacich (1972) demonstrated that there is always a positive eigenvalue λ such that all elements of the corresponding eigenvector are greater than or equal to zero.

The Bonacich index of centrality of a point i is therefore defined as:

$$C_{\mathrm{Bon}}(i) = s_i, \tag{9}$$

where the eigenvector $\boldsymbol{S}$ corresponds to λ_1, the largest eigenvalue, and is normalized to a length of 1, i.e.,

$$|\boldsymbol{S}| = \sqrt{\sum_i s_i^2} = 1. \tag{10}$$

We also define, as an index of centrality for future reference in following sections, the quantity:

$$C_{\lambda}(i) = \lambda_1, \tag{11}$$

which is the expansion of the system along the eigenvector $\boldsymbol{S}$.

For the network of Fig. 1, C_{Bon} index does discriminate points 41 and 42 from point 45 with $C_{\mathrm{Bon}}(41–42) = 0.358$ and $C_{\mathrm{Bon}}(45) = 0.427$ (cf. Table 1). Point 45 is considered more central by C_{Bon}, as was the case with C'_{Inf}. In the case of the AIDS network (Fig. 2), individual 16 is ranked the most ‘‘popular’’. Individuals 22, 11 and 20 are ranked more popular than individual 26, which is coherent with the spirit of this measure, mainly because they are related to the most popular individual.

Here again, one has to manipulate matrices to obtain the Bonacich centrality index. Even with specialized algorithm (we only need the eigenvector corresponding to the largest eigenvalue), the process is very time- and memory-consuming for large networks (say N greater than a few thousands of points). Moreover, the indices are normalized and no information is available to allow comparison of centrality scores between points from different components.

4. Mapping-based centrality indices

In this section, we present centrality indices based on the iteration of simple mappings. These indices consider all the paths of a connected network. Some of them will be constructed to allow comparison of points between different components of variable size. One of the mappings presented is shown to converge rapidly, considerably reducing computer time, even for large networks.

4.1. Normalized Bonacich mapping

The Bonacich index of centrality, presented in Section 3.4, can also be obtained without having to solve the eigensystem (Eq. 8). A simple measure of popularity is just the number of contacts each person has (i.e., the degree). Bonacich further argued that second-order indices can be obtained by weighting each contact by the number of contacts they have. Then in a recursive manner, if second-order indices are better than first-order ones, a third-order index, where the contacts of each individual are weighted by their second-order indices, should be even better, and so on.

Let the vector $\boldsymbol{S}^{0\prime}$ denote the popularity score of individuals at the initial stage of this iterating process. First-order indices, i.e., after one iteration, are given by $\boldsymbol{S}^{1\prime} = \mathbf{A}\boldsymbol{S}^{0\prime}$, where $\mathbf{A}$ is the adjacency matrix describing the contact network. Second-order indices are given by $\boldsymbol{S}^{2\prime} = \mathbf{A}\boldsymbol{S}^{1\prime}$ and so on, such that nth-order indices are given by $\boldsymbol{S}^{n\prime} = \mathbf{A}\boldsymbol{S}^{n-1\prime}$.

An example of this process is illustrated in Table 3 for component L of Fig. 1. Here, the points in component L can be seen as the individuals that we want to characterize according to their popularity level. The variable $s_i^{n\prime}$ denotes the nth order popularity score of individual i. Note that from order 2 onwards, individual 45 can be discriminated from individuals 41 and 42 (which are structurally equivalent). Thus, based on the Bonacich sequential scheme, individual 45 is more central than individuals 41 and 42 because of higher-order popularity scores.

As n increases, popularity scores tend to infinity. However, Bonacich showed that the process converges to a finite set of nonzero popularity indices if the result of each iteration step is divided by λ_1, the largest eigenvalue of $\mathbf{A}$. The iterating procedure is then:

$$\boldsymbol{S}^n = \mathbf{A}\boldsymbol{S}^{n-1}/\lambda_1. \tag{12}$$

Table 3
Dynamical evolution of the Bonacich mapping applied to component L of Fig. 1. For each iteration step n, $s_i^{n\prime}$ and s_i^n are the popularity score and the normalized popularity score of point i, respectively

n/i	$s_i^{n\prime}$			s_i^n		
	41 and 42	43 and 44	45	41 and 42	43 and 44	45
0	1	1	1	1.000	1.000	1.000
1	2	3	2	0.365	0.548	0.365
2	5	7	6	0.369	0.516	0.442
3	12	18	14	0.357	0.535	0.416
4	30	44	36	0.359	0.527	0.431
5	74	110	88	0.357	0.531	0.425
6	184	272	220	0.358	0.529	0.428
7	456	676	544	0.358	0.530	0.427
8	1132	1676	1352	0.358	0.530	0.427
9	2808	4160	3352	0.358	0.530	0.427
10	6968	10320	8320	0.358	0.530	0.427

At the limit (when $n \to \infty$), the sequence converges and one obtains $\boldsymbol{S}^n = \boldsymbol{S}^{n-1}$ (with $\boldsymbol{S}^n \to \boldsymbol{S}$). Bonacich mapping (Eq. 12) then becomes:

$$(\mathbf{A} - \lambda_1 \mathbf{I})\boldsymbol{S} = 0,$$

the eigensystem (Eq. 8) presented in Section 3.4. In the iterative process (Eq. 12), the convergence to the eigenvector requires a sufficiently good estimate of λ_1, the largest eigenvalue; otherwise, the process goes to zero or infinity.

In order to improve numerical stability, instead of dividing by λ_1, we propose to normalize the popularity vector at each iteration step. This does not change the dynamical property of the mapping $\boldsymbol{S}^n = \mathbf{A}\boldsymbol{S}^{n-1}$, for which the vector $\boldsymbol{S}$ tends to get aligned along the eigenvector corresponding to the largest eigenvalue as n increases. That is, the vector $\boldsymbol{S}$ is simply rescaled at each iteration step and the normalized Bonacich mapping can be rewritten as:

$$\boldsymbol{S}^n = \mathbf{A}\boldsymbol{S}^{n-1} / |\mathbf{A}\boldsymbol{S}^{n-1}|, \tag{13}$$

where $|\mathbf{A}\boldsymbol{S}^{n-1}|$ is the norm (modulus) of the vector $\boldsymbol{AS}^{n-1}$.

In Table 3, the normalized popularity scores, denoted s_i^n, for component L of the network of Fig. 1 are presented. One can see that the normalized sequence converges rapidly to the eigenvector already obtained in Section 3.4 (cf. C_{Bon} in Table 1). Unfortunately, this is not always the case. Consider component C of the same network, for which the first iteration steps of Bonacich's mappings are presented in the left-hand side of Table 4. It can be seen that the process is trapped into a periodical orbit (of period 2) and thus could never converge to the eigenvector. As one could imagine, this ''pathology'' is certainly not exclusive to this case.

The left-hand side of Table 5 presents the results (for few individuals) of Bonacich's mapping for the AIDS network of Fig. 2. Rows beginning with Rn in Table 5

Table 4
Comparison of the dynamical evolution of the Bonacich mapping and the cumulative nomination mapping applied to component C in Fig. 1
For each iteration step n, $s_i^{n\prime}$, s_i^n, $p_i^{n\prime}$ and p_i^n are the popularity score, the normalized popularity score, the cumulative number of nominations, and the normalized cumulative number of nominations of point i, respectively. Here, the cumulative nomination mapping is normalized the same way as the normalized Bonacich mapping (see Section 4.1)

n / i	Bonacich mapping				Cumulative nomination mapping			
	$s_i^{n\prime}$		s_i^n		$p_i^{n\prime}$		p_i^n	
	4 and 6	5	4 and 6	5	4 and 6	5	4 and 6	5
0	1	1	1.000	1.000	1	1	1.000	1.000
1	1	2	0.408	0.816	2	3	0.485	0.728
2	2	2	0.577	0.577	5	7	0.503	0.704
3	2	4	0.408	0.816	12	17	0.500	0.708
4	4	4	0.577	0.577	29	41	0.500	0.707
5	4	8	0.408	0.816	70	99	0.500	0.707
6	8	8	0.577	0.577	169	239	0.500	0.707
7	8	16	0.408	0.816	408	577	0.500	0.707
8	16	16	0.577	0.577	985	1393	0.500	0.707
9	16	32	0.408	0.816	2378	3363	0.500	0.707
10	32	32	0.577	0.577	5741	8119	0.500	0.707

correspond to the growth rate of the popularity scores between iteration step $n-1$ and n before renormalization, i.e., the quantity:

$$g_i^n = \left(a_{i1} s_1^{n-1} + \ldots + a_{iN} s_N^{n-1}\right) / s_i^{n-1}. \tag{14}$$

It has been argued previously that the popularity scores tend to get aligned along the eigenvector corresponding to λ_1, the largest eigenvalue. As can be seen, the values g_i^n converge appropriately to $\lambda_1 = 3.117$ (cf. C_λ in Table 2) as the popularity vector $\boldsymbol{S}$ get aligned along the direction of the eigenvector. However, one has to perform more than 10,000 iterations, dandling between two states while converging slowly to the solution. Bonacich mapping may therefore not always be the most ideal method to obtain centrality indices.

4.2. Cumulated nomination mapping

Let us imagine a procedure by which the popularity of individuals, within a contact network, is obtained by cumulating nominations. Initially (stage 0 of the process), every individual gets one nomination. Then, after the first round of nominations (stage 1), each individual gets additional nominations from their contacts, weighted by the number of nominations their contacts already have, which is 1 at this stage. Thus, a contact with many nominations will be considered more important than a contact with only few nominations. The process is repeated such that individual cumulates nominations every new round. Therefore, at each new round, the new nominations that an individual

Table 5
Comparaison of the dynamical evolution of the normalized Bonacich mapping and the cumulative nomination mapping applied to the AIDS network of Fig. 2 for selected individuals
For each iteration step n, s_i^n and p_i^n are the normalized popularity score and the proportion of the cumulative number of nominations of individual i, respectively. Rows beginning with Rn list the expansion of the corresponding quantities from step $n-1$ to step n before renormalization (see text).

n/i	Bonacich mapping (s_i^n)					Cumulative nomination mapping (p_i^n)				
	1	2	5	11	16	1	2	5	11	16
0	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1	0.063	0.125	0.314	0.188	0.502	0.656	0.984	1.967	1.311	2.951
2	0.042	0.127	0.170	0.340	0.340	0.437	0.961	1.659	2.009	2.882
3	0.042	0.070	0.259	0.210	0.587	0.368	0.805	1.770	2.023	3.609
4	0.023	0.098	0.146	0.328	0.401	0.297	0.744	1.634	2.169	3.728
5	0.032	0.055	0.220	0.209	0.629	0.262	0.674	1.604	2.169	4.066
R6	1.721	4.608	1.846	4.736	2.127	3.570	3.767	3.829	4.100	4.160
11	0.022	0.043	0.180	0.214	0.669	0.173	0.513	1.369	2.295	4.813
12	0.014	0.065	0.120	0.314	0.465	0.168	0.503	1.356	2.310	4.879
13	0.021	0.043	0.177	0.216	0.673	0.164	0.495	1.346	2.323	4.936
14	0.014	0.063	0.120	0.314	0.469	0.161	0.490	1.339	2.336	4.983
15	0.020	0.043	0.176	0.217	0.675	0.159	0.486	1.335	2.347	5.024
R16	2.107	4.573	2.136	4.512	2.175	4.058	4.077	4.095	4.120	4.130
101	0.020	0.044	0.175	0.223	0.677	0.156	0.486	1.358	2.441	5.252
102	0.014	0.063	0.124	0.315	0.479	0.156	0.486	1.358	2.441	5.252
103	0.020	0.044	0.175	0.223	0.677	0.156	0.486	1.358	2.441	5.252
104	0.014	0.063	0.124	0.315	0.479	0.156	0.486	1.358	2.441	5.252
105	0.020	0.044	0.175	0.223	0.677	0.156	0.486	1.358	2.441	5.252
R106	2.204	4.408	2.205	4.406	2.206	4.117	4.117	4.117	4.117	4.117
1001	0.019	0.047	0.170	0.236	0.656					
1002	0.015	0.061	0.131	0.305	0.507					
1003	0.019	0.047	0.170	0.236	0.656					
1004	0.015	0.061	0.131	0.305	0.507					
1005	0.019	0.047	0.170	0.236	0.656					
R1006	2.409	4.033	2.410	4.032	2.411					
10,001	0.018	0.054	0.153	0.271	0.591					
10,002	0.017	0.055	0.151	0.275	0.583					
10,003	0.018	0.054	0.153	0.271	0.591					
10,004	0.017	0.055	0.151	0.275	0.583					
10,005	0.018	0.054	0.153	0.271	0.591					
R10,006	3.075	3.161	3.075	3.161	3.075					
100,001	0.017	0.054	0.152	0.273	0.587					
100,002	0.017	0.054	0.152	0.273	0.587					
100,003	0.017	0.054	0.152	0.273	0.587					
100,004	0.017	0.054	0.152	0.273	0.587					
100,005	0.017	0.054	0.152	0.273	0.587					
R10,0006	3.117	3.117	3.117	3.117	3.117					

obtains from his contacts are added to those he already has. After a while, individuals are ordered in the function of their cumulated number of nominations; more central individuals having cumulated more nominations.

An example of this process is illustrated in Table 6 for component L of Fig. 1 where points represent nominees (individuals). The variable $p_i^{n\prime}$ denotes the number of nominations cumulated by individual i after n stages. In the initial stage (stage 0), everybody obtains one nomination. In step 1, individual 41 whose contacts are 42 and 44, cumulates three nominations, i.e., one nomination in step 0 and one nomination for each of his two partners in step 0. In step 2, the cumulative number of nominations of individual 41 increases to $3+3+4=10$, i.e., three from himself in step 1, three from his contact 42 and four from his contact 44. In step 3, we obtain $10+10+14=34$ nominations, and so on. In Table 6, we see that after two iterations, individual 45 can be discriminated from individuals 41 and 42 (which are structurally equivalent). Thus, based on cumulated nominations mapping, individual 45 is more central than individuals 41 and 42 because he cumulates more nominations.

Although our cumulated nomination scheme seems very similar to the one proposed by Bonacich (1972) (see Section 4.1.), it presents some differences and advantages.

Mathematically, the cumulated nomination scheme is expressed by the following mapping:

$$p_i^{n+1\prime} = p_i^{n\prime} + \sum_j a_{ij} p_j^{n\prime}, \tag{15}$$

where $p_i^{n\prime}$ is the cumulative number of nominations for individual i in step n with $p_i^{0\prime}=1$, and a_{ij} are entries of the adjacency matrix of the contact network. In order to define measures of manageable size and to assess their convergence properties, proportions will be used to characterize the centrality of a point in a network. The quantities $p_i^{n\prime}$ are then normalized to obtain the proportions:

$$p_i^n = \frac{p_i^{n\prime}}{\sum_j p_j^{n\prime}}, \tag{16}$$

in step n. Since it is equivalent and computationally more convenient, proportions p_i^n are iterated and normalized at each step. The mapping (Eq. 15) then becomes:

$$p_i^{n+1} = \frac{p_i^n + \sum_j a_{ij} p_j^n}{\sum_k \left[p_k^n + \sum_j a_{kj} p_j^n \right]}, \tag{17}$$

with of course:

$$\sum_i p_i^n = 1. \tag{18}$$

Table 6 shows the evolution of the proportions p_i^n when the mapping (Eq. 17) is applied to the component L in Fig. 1 for the first iteration steps. It can be seen that these quantities converge to a fixed value as n increases, i.e., the system reaches an equilibrium. Theoretically, these values represent the attractive fixed point of a dynami-

Table 6
Dynamical evolution of the cumulated nominations mapping applied to component L of Fig. 1
For each iteration step n, $p_i^{n\prime}$, p_i^n and $g_i^{n\prime}$ are the cumulative number of nominations, the proportion of the cumulative number of nominations, and the nomination growth rate of point i, respectively.

n/i	$p_i^{n\prime}$			p_i^n			$g_i^{n\prime}$		
	41 and 42	43 and 44	45	41 and 42	43 and 44	45	41 and 42	43 and 44	45
0	1	1	1	0.2000	0.2000	0.2000	–	–	–
1	3	4	3	0.1765	0.2353	0.1765	3.000	4.000	3.000
2	10	14	11	0.1695	0.2373	0.1864	3.333	3.500	3.667
3	34	49	39	0.1659	0.2390	0.1902	3.400	3.500	3.545
4	117	171	137	0.1641	0.2398	0.1921	3.441	3.490	3.513
5	405	596	479	0.1632	0.2402	0.1931	3.462	3.485	3.496
6	1406	2076	1671	0.1628	0.2404	0.1935	3.472	3.483	3.489
7	4888	7229	5823	0.1626	0.2405	0.1937	3.477	3.482	3.485
8	17,005	25,169	20,281	0.1625	0.2406	0.1938	3.479	3.482	3.483
9	59,179	87,624	70,619	0.1625	0.2406	0.1939	3.480	3.481	3.482
10	205,982	305,046	245,867	0.1625	0.2406	0.1939	3.481	3.481	3.482

cal system (in five dimensions) (see, e.g., Devaney, 1989). Without going into extensive details, note that, for a mapping $x^{n+1}=f(x^n)$, a fixed point is a point x^* such that $f(x^*)=x^*$. A fixed point x^* is attractive if the derivative of f, evaluated at the fixed point x^*, lies between -1 and 1 (more generally, for a multi-dimensional system, the eigenvalues of the Jacobian of the mapping evaluated at the fixed point must all lie between -1 and 1). Table 6 shows the convergence of the system to the attractive fixed point. In Fig. 3, the convergence is illustrated using three different sets of initial conditions corresponding to values of p_i^0 which are: (1) chosen randomly in the interval [0,1]; (2) all null except for $p_{41}^0=1$; or (3) proportional to the degree of the point (and scaled such that $\sum_i p_i^0=1$). As suggested by Fig. 3, there is only one asymptotic solution to the mapping of Eq. 17.

Analogous to the Bonacich mapping, the convergence of the method is insured because the cumulated nominations mapping (Eq. 15) is a linear mapping that converges to the eigenvector corresponding to the largest eigenvalue of the matrix of the linear system. This vector is simply rescaled from step to step and one obtains the mapping (Eq. 17). To see how Bonacich's and our mappings are related, let us express the cumulative nomination mapping in matrix form as:

$$\boldsymbol{P}^{n+1\prime}=\boldsymbol{P}^{n\prime}+\mathbf{A}\boldsymbol{P}^{n\prime}=(\mathbf{A}+1)\boldsymbol{P}^{n\prime},$$

where $\boldsymbol{P}^{n\prime}$ is a vector containing the cumulated number of nominations in stage n. The corresponding eigensystem can be written:

$$[\mathbf{A}-(\lambda'-1)\mathbf{I}]\boldsymbol{P}'=0,$$

which is completely equivalent to the Bonacich case (see Eq. 8) with the transformation:

$$\lambda=\lambda'-1.$$

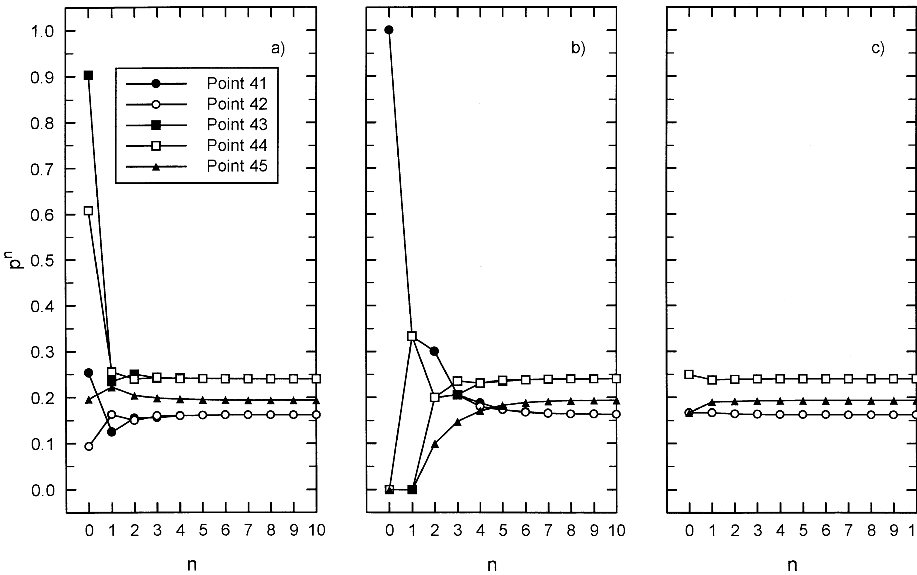

Fig. 3. Convergence of the cumulated nomination centrality indices for component L in Fig. 1 (see text).

In Table 6, one can see that the growth rate (compare with Eq. 14):

$$g_i^{n\prime} = \left(p_i^{n-1} + a_{i1} p_1^{n-1} + \ldots + a_{iN} p_N^{n-1} \right) / p_i^{n-1}, \tag{19}$$

converges appropriately to $\lambda' = \lambda + 1 = C_\lambda + 1 = 3.48$ (see Eq. 11 and Table 1) for each i.

Thus, up to a normalization constant, the cumulated nomination mapping (Eq. 17) converges exactly to the same solution as that of Bonacich mapping (Eq. 13) except that it does not suffer from the same bimodal convergence pathology as Bonacich's mapping does (see Section 4.1). Instead of having, as is the case with Bonacich's mapping, the load of nominations exchanged between contacts at each step, the cumulative nomination mapping builds on the total nominations that have been cumulated so far. Table 4 shows the first iteration steps of the cumulative nomination mapping (Eq. 17), with normed p_i^n to allow comparison with Bonacich mapping. One can see that the cumulative nomination mapping converges rapidly to the adequate eigenvector (cf. C_{Bon} in Table 1 for component C).

For a connected network (or within a component), the mapping (Eq. 17) can be used to define a global centrality index to characterize each point of the network. Moreover, since the scaling of the indices is arbitrary (the values p_i^n sum to 1, see Eqs. 17 and 18), it is possible to define indices that do not depend on the size of the network. In Fig. 1, every point within any of the components A, B, D, F, H, J and M (circle or complete graphs) is equivalent to any other. Ideally, a centrality index should emphasize this. In other words, a structurally averaged point should get an average centrality score, independently of the size of its component.

A new measure of centrality, the cumulated nominations index of centrality $C_{CN}(i)$, which can be used to characterized point i within a connected network (or within a connected component of a network), is thus defined by:

$$C_{CN}(i) = N \lim_{n \to \infty} p_i^n, \tag{20}$$

where p_i^n is obtained from the mapping (Eq. 17) with the initial condition $p_i^0 = 1$. The multiplicative factor N insures that a point with an average centrality will obtain a score of 1, independently of N.

In practice, the iterative process (Eq. 20) can be stopped once the system has converged to the attractive fixed point, within a predefined interval of precision. For example, iterations could be stopped once all the differences d_i^n, in absolute value, between p_i^n and p_i^{n-1} are inferior to a stopping criterion. More formally, we choose to stop the iterating process when $\max_i\{|p_i^n - p_i^{n-1}|\} < \varepsilon$ with $\varepsilon \ll 1$ (recall that the values p_i^n are already scaled with respect to 1, see Eq. 18).

The cumulated nominations centrality indices C_{CN} are given in Table 1 for the network of Fig. 1. Within each component, points that are more central than average obtain centrality scores above 1. In component L, point 45 gets a score very close to the average with $C_{CN}(45) = 0.970$, while points 41 and 42 can be considered less central with $C_{CN}(41) = C_{CN}(42) = 0.812$. This result is in contradiction with the one based on S–Z measure, C_{Inf}, but coherent with that obtained from Altmann measure, C'_{Inf} (cf. Section 3.3). Moreover, results obtained from C_{CN} are certainly coherent with those obtained from Bonacich's index, C_{Bon}, since both sets of indices are the same (within a component), up to a scaling factor. This is clearly seen from Table 2 within which individuals forming the AIDS network are classified the same way by C_{Bon} or C_{CN}.

4.3. Mapping-based centrality indices for networks with many components

Let us focus our attention on the quantity λ' which has just been shown to be the limit of $g_i^{n\prime}$ when n goes to infinity (see Eq. 19). It represents the growth rate of the number of nominations at equilibrium. Within a component of a network, every point gets the same growth rate. This growth rate, which is a quantity related to the limiting number of new nominations at each step within a component, therefore characterizes each point within a same component.

Thus, in order to compare points between different components, a new measure of centrality is introduced. It is defined by:

$$C_{CNGr}(i) = \lambda' = \lim_{n \to \infty} g_i^{n\prime} = \lim_{n \to \infty} \left[\frac{p_i^n + \sum_j a_{ij} p_j^n}{p_i^n} \right] = \lim_{n \to \infty} \left[1 + \frac{\sum_j a_{ij} p_j^n}{p_i^n} \right], \tag{21}$$

where p_i^n is obtained from the mapping (Eq. 17) with the initial condition $p_i^0 = 1$. Cumulated nominations growth rate centrality scores, C_{CNGr}, are given in Table 1 for the network of Fig. 1. As expected, the cumulative nomination growth rate is the same for any point within the same component and is related to the largest eigenvalue of the contact matrix, $C_\lambda(i)$, by $C_{\mathrm{CNGr}}(i) = C_\lambda(i) + 1$. This is also verified for the AIDS network of Fig. 2, for which we have $C_{\mathrm{CNGr}}(i) = C_\lambda(i) + 1 = 3.117 + 1 = 4.117$ (see caption of Table 2).

More importantly, the cumulative nomination growth rate centrality index allows the comparison of different components by adjusting the cumulated nominations centrality index, C_{CN}, for the level of nomination activity within a component. Along these lines, a new multi-component cumulated nominations centrality index, $C'_{\mathrm{CN}}(i)$, characterizing a point i located in a network made of one or more components, is defined by:

$$C'_{\mathrm{CN}}(i) = C_{\mathrm{CN}}(i) \times C_{\mathrm{CNGr}}(i). \tag{22}$$

With this measure, the more central a point is within its component, the larger its centrality score will be *and* the higher the frequency of nominations is within its component, the larger its centrality score will also be. Multi-component cumulated nominations centrality indices are given in Table 1 for the network of Fig. 1. Despite an equal cumulated nominations centrality index (C_{CN}) of one, points in complete components, such as H or M, are rated by C'_{CN} as being more central than those in cyclic components, F or J, because of a higher within-component nominations activity.

Moreover, points in large components should be considered more central than others; otherwise, structurally equivalent, in smaller components, simply because more points are reachable within large components, resulting in higher potential impact and thus, a higher centrality (cf. discussion leading to the definition of C_{CSize} in Section 3.2). This can easily be taken into account by defining, for a network made of one or more components, the corrected multi-component cumulated nominations centrality index for the point i as:

$$C''_{\mathrm{CN}}(i) = C_{\mathrm{CN}}(i) \times C_{\mathrm{CNGr}}(i) \times C_{\mathrm{CSize}}(i) = C'_{\mathrm{CN}}(i) \times C_{\mathrm{CSize}}(i), \tag{23}$$

where $C_{\mathrm{CSize}}(i)$ corresponds to the relative component's size of point i. Results for Fig. 1 are presented in Table 1. For example, points located in cyclic components D, F and J are neither discriminated by C_{CN} nor by C'_{CN} but are by C''_{CN}. For the one-component AIDS network in Fig. 2, $C_{\mathrm{CSize}}(i) = 1$ and $C_{\mathrm{CN}}(i) = C'_{\mathrm{CN}}(i)$ for all i.

4.4. Computational advantages of mapping-based indices for large networks

It has previously been argued that methods to obtain centrality indices based on matrix manipulations quickly become CPU and memory-intensive as the size of the network increases. Despite the use of efficient sparse matrix manipulation algorithms, Altmann (1993) obtained prohibitive computer time (more than 17 h on a VAXStation 3100) to calculate information centrality indices (C_{Inf} or C'_{Inf}), for random networks of moderate size ($N = 3000$).

Methods based on mappings are therefore very interesting because they are easy to implement and converge rapidly to the desired centrality indices. It is also very easy to

Table 7
Linked list notation vs. adjacency matrix notation to describe component L of the contact network in Fig. 1

Adjacency matrix notation						i	Linked list notation	
i/j	41	42	43	44	45		$\|L_i\|$ (number of links)	L_i (links)
41	0	1	0	1	0	41	2	{42,44}
42	1	0	1	0	0	42	2	{41,43}
43	0	1	0	1	1	43	3	{42,44,45}
44	1	0	1	0	1	44	3	{41,43,45}
45	0	0	1	1	0	45	2	{43,44}

express the mapping in such a way to avoid any matrix manipulations. Let L_i denote the set of points j that are adjacent to i, i.e.,

$$L_i = \{ j | a_{ij} = 1 \}.$$

The mapping (Eq. 17) then becomes:

$$p_i^{n+1} = \frac{p_i^n + \sum_{j \in L_i} p_j^n}{\sum_k \left[p_k^n + \sum_{j \in L_k} p_j^n \right]}. \tag{24}$$

Table 7 compares adjacency matrix to linked list notation for component L of the network in Fig. 2. Linked lists such as the one shown in Table 7 are easy to implement and manage with memory-efficient modern programming language such as C or C ++ (or any language that allows dynamic memory allocation).

For networks with degrees that do not increase with N (i.e., the number of links are of the order N), the memory and the CPU time required to load the complete contact network and to perform one iteration of the mapping, (Eq. 24), are of the order N.

To give an idea of the computational performance offered by mapping-based methods, the cumulative nomination centrality indices converge within seconds for simulated networks of moderate size ($N \sim 5000$) and within minutes for simulated networks of larger size ($N \sim 50{,}000$). Those are approximative benchmarks for typical Unix workstations clocked at 200 MHz with enough memory (RAM, not swap space) to load the complete network.

5. Application and validation of structural measures (SMs)

In this section, we examine and discuss the properties of the different measures (new and classic ones) defined in the previous sections, in order to characterize the structure of social networks. The centrality measures are compared by first considering the case of a connected network (one component) and then the case of multi-component networks (many components). Finally, questions regarding graph comparability and scaling laws for SMs are studied in the light of generated (simulated) networks of variable sizes.

5.1. Comparative analysis of SMs

5.1.1. Connected networks

Let us go back to the contact network of 40 homosexual men with AIDS (cf. Fig. 2), used as an illustrative example in previous sections, and go a further in the discussion and comparison of the different SMs introduced. Stephenson and Zelen (1989) have already studied the structure of this network using their information centrality index, C_{Inf} (see Section 3.3). The results of a similar analysis, which compares the various centrality indices described in this paper, are presented in Table 2. For each measure, the ranks of individuals are listed in ascending order of centrality from the most central (rank 1) to the less central (rank 40). For each measure, the index value corresponding to each rank is presented in parenthesis. Remember that, for a connected network (cf. Section 4.2), C_{Bon}, C_{CN} and C'_{CN} $(= C_{\mathrm{CN}} \times C_{\mathrm{CNGr}} = C_{\mathrm{CN}} \times 4.117)$ are equivalent up to a normalization constant. Thus, the ranks are exactly the same for C_{Bon}, C_{CN} and C'_{CN}. However, we will see that, with regards to their interpretation, C_{CN} or C'_{CN} are still preferred over C_{Bon}.

At first glance, the network shown in Fig. 2 suggests that individuals forming these three distinct groups {12,15,17,21}, {3,4,6} and {39,40} should get similar centrality scores since they are structurally equivalent. The results produced in Table 2 show that it is indeed the case and that all these centrality measures are valid to identify homogeneous individuals who are symmetric with respect to structural characteristics. However, not all measures have the same discriminant power to detect more subtle structural heterogeneities between individuals. Consider individuals 1 and 10 who both get $C_{\mathrm{Deg}} = 1$ and $C_{\mathrm{Ey}} = 0.256$. A careful examination of Fig. 2 suggests that these individuals are similar but not structurally equivalent. They are both located at a distance 3 from the ‘‘pivotal’’ individual 11, leading to the same eccentricity score.

Individuals 1 and 10 have one contact with individuals 2 and 9, respectively (with $C_{\mathrm{Deg}}(2) = C_{\mathrm{Deg}}(9) = 2$), who themselves have one contact, with individuals 5 and 8, respectively, who are both directly in contact with individual 11. However, individual 5 ($C_{\mathrm{Deg}}(5) = 5$) is more central than individual 8 ($C_{\mathrm{Deg}}(8) = 3$) as indicated by closeness indices ($C_{\mathrm{CS}}(1) = 0.169$ and $C_{\mathrm{CS}}(10) = 0.166$). Note that the other centrality indices also rank individual 1 more centrally than individual 10 (see Table 2).

Closeness is slightly better than eccentricity but still attributes the same score of 0.262 to individuals 12 and 28, yet, Fig. 2 shows that they are not structurally equivalent. Individual 12 only has partner 16 who happens to be the most central individual according to intuition and most measures. Comparatively, individual 28 has three partners who are more peripheral than individual 16. The remaining centrality indices discriminate between individuals 12 and 28. They are, respectively, ranked 10th (0.299) and 6th (0.348) by C_{Inf}, 13th (0.329) and 5th (0.418) by C'_{Inf}, and in reverse order, 7th (1.685) and 12th (1.185) by C_{CN} (or C_{Bon} or C'_{CN}). If we assume that the detection of structural heterogeneity is a positive result and that of homogeneity is a negative one, then, to borrow some epidemiological terminology (Kleinbaum et al., 1982), we observe that all centrality measures have a good specificity but only a few have a good sensitivity. If we pursue the issue further, one can ask which of these two individuals, 12 or 28, is the most central? The answer to this question is subject to interpretation and depends on the specific objectives of the study undertaken.

As suggested by Altmann (1993), more than one measure should be used to study a given network. What is of upmost importance is to understand the meaning of the measures and to be able to choose those most likely to correlate with the phenomenological process of interest. For example, the cumulated nominations growth rate, C_{CNGr}, is used in Section 4.3 to define a multi-component cumulated nominations centrality index (C'_{CN}) that can be used to compare results between different independent networks (or components) of different sizes. For the network in Fig. 2, the C_{CNGr} index is 4.117 per cycle (cyc^{-1}). According to our cumulated nominations scheme, an average individual (individual 8 is close to this) in the AIDS example gets nominated (or possibly risk of being exposed to HIV) at a rate of $1.000 \times 4.117 = 04.117\ \mathrm{cyc}^{-1}$ compared to a rate of $5.252 \times 4.117 = 21.62\ \mathrm{cyc}^{-1}$ for the most central individual 16. The reason why the value of C'_{CN} has a direct and convenient interpretation is due to the fact that C_{CN} indices are normalized such that they represent the proportion of the cumulated number of nominations received. This information could also be correlated with some cumulative exposure to HIV, for example. Bonacich's indices, C_{Bon}, give (up to a normalization constant) information similar to that of C_{CN} on the relative positions of individuals within networks made of one component. However, the value of the indices, in absolute term, does not have any specific, concrete or direct interpretation (as opposed to C_{CN}) because they are standardized such that the modulus (the length) of the eigenvector equals 1.

Table 8 shows the Pearson correlation matrix between pairs of the various centrality indices presented in Table 2. As expected, one can immediately see the perfect correlations between C_{Bon}, C_{CN} and C'_{CN} indices. Since the AIDS network is made of a single component, those indices are the same up to a scaling factor (see Sections 4.2 and 4.3). The eccentricity index, C_{Ey}, correlates negatively with the others since, of course, the most ''eccentric'' (peripheral) individual is less central. The degree index, C_{Deg},

Table 8
Pearson correlation matrix between various centrality indices applied to the AIDS network of Fig. 2 (see Table 2)

	C_{Deg}	C_{Ey}	C_{Cs}	C_{Inf}	C'_{Inf}	C_{Bon}	C_{CN}	C'_{CN}
C_{Deg}	1.000	−0.436	0.658	0.678	0.801	0.727	0.727	0.727
C_{Ey}	−0.436	1.000	−0.902	−0.880	−0.805	−0.703	−0.703	−0.703
C_{Cs}	0.658	−0.902	1.000	0.978	0.938	0.902	0.902	0.902
C_{Inf}	0.678	−0.880	0.978	1.000	0.972	0.882	0.882	0.882
C'_{Inf}	0.801	−0.805	0.938	0.972	1.000	0.864	0.864	0.864
C_{Bon}	0.727	−0.703	0.902	0.882	0.864	1.000	1.000	1.000
C_{CN}	0.727	−0.703	0.902	0.882	0.864	1.000	1.000	1.000
C'_{CN}	0.727	−0.703	0.902	0.882	0.864	1.000	1.000	1.000

correlates weakly with other indices. This empasizes the local nature of this measure and the fact that it does not capture the global structure of a network, as other measures do. More interestingly, the high correlations of 0.978 between C_{Inf} and C_{Cs} and of 0.938 between C'_{Inf} and C_{CS} suggest that these measures (or information transmitted between individuals) seem greatly influenced by the proximity, or closeness, of individuals within the network (which make sense). This is also verified, although to a lesser extent, by the correlation of 0.902 between C_{Bon}, C_{CN} or C'_{CN} and C_{Cs}. These latter measures, which are coherent with closeness with respect to centrality, are sufficiently different to be considered relevant.

5.1.2. Multi-component networks

One goal of this paper is to identify centrality indices that are suitable to characterize the structure of networks made of many components and various sizes. We have developed our centrality indices with these properties in mind (i.e., C'_{CN} and C''_{CN}). In our analyses, all other indices are standardized by the size of the network (e.g., C_{CSize}, C_{Ey} and C_{Cs}).

Centrality indices are given in Table 1 for the multi-component network of Fig. 1. The component size, C_{CSize}, and the degree, C_{Deg}, provide some information on the importance of a given point within a network. Moreover, their meaning facilitates the interpretation and comparison of points between different components. Since C_{CSize} cannot distinguish points within a component and since C_{Deg} is exclusively local, a combination of both would not distinguish structurally different points with the same degree within a component. The eccentricity, C_{Ey}, and the closeness, C_{Cs}, better reflect the structure within a component but are standardized to one (the largest possible value of the index for a network of a given size) for convenience rather than helpful interpretation. The indices C_{Inf} and C'_{Inf} are interpreted as the harmonic and the arithmetic average information that can be transmitted between a point and any other reachable point, respectively. While these indices permit to discriminate between points within a component, the relation between each index value and the size of their component is not clear. For example, information centrality scores of each point in cyclic components D, F and J in Fig. 1 decrease as N increases with $C_{\mathrm{Inf}} = 2.250$, $C_{\mathrm{Inf}} = 1.600$, $C_{\mathrm{Inf}} = 1.250$ and $C'_{\mathrm{Inf}} = 1.000$, $C'_{\mathrm{Inf}} = 0.917$, $C'_{\mathrm{Inf}} = 0.833$ for D,F and J, respectively. Intuitively, it makes sense that less information gets transmitted between points as they become more distant when the size of the component increases but no one knows how. Altmann (1993) suggested that averages (4) and (5) should be rescaled by $\sqrt{N}$ instead of N for planar networks which are uncommon in social networks (see Section 3.3).

Our cumulated nominations centrality index C_{CN}, discriminates points within a component. The nomination growth rate, C_{CNGr}, corrects this standardized index to weight the importance of a point relative to the nomination activity within the component (see Eq. 22). Finally, the index C''_{CN} corrects C'_{CN} to take into account the relative size of the component (see Eq. 23). The relation between these indices is clear as well as their interpretation. A structurally average point would yield $C_{\mathrm{CN}} = 1$, with C'_{CN} being the nomination activity within its component. Another point within the same component, being nominated half the average, would yield half the score. Two average points, from two different components, would yield $C_{CN} = 1$ but different C'_{CN} and/or C''_{CN}, depending on the activity and the size of their corresponding component. For example, in network of Fig. 1, despite an equal cumulated nominations centrality index $C_{CN} = 1$, points within complete components H or M are rated by C'_{CN} to be more central than those within cyclic components, F or J, because of a higher within-component nominations activity. Points located in cyclic components D, F and J are discriminated neither by C_{CN} nor by C'_{CN} but are by C''_{CN} because of the different components' relative size.

Table 9 shows the Pearson correlation matrix between various centrality indices for the network in Fig. 1. Note that C_{Bon}, C_{CN}, C'_{CN} and C''_{CN} no longer correlate perfectly. The eigenvalue C_{λ} and our nominations growth rate C_{CNGr} correlate perfectly since they are related by $C_{CNGr} = C_{\lambda} + 1$. Bonacich's index ($C_{Bon}$) correlates negatively with the relative size of the components, C_{CSize}, while our index (C_{CN}) does not correlate at all. Remember that our index, C_{CN}, is standardized such that an average point gets a score of 1, independently of the size of its component (connected network). Surprisingly, the information-based indices, C_{Inf} and C'_{Inf}, correlate positively with CSize, suggesting that more information is transmitted in larger components. In this case, this can be simply explained by the fact that larger components produce, on average, higher degrees scores as illustrated by the positive correlation between C_{CSize} and C_{Deg}. Our index C'_{CN} correlates highly with C_{Deg} (0.982) wrongly suggesting, as will be seen in Section 5.2, a limited use for our measure.

To study correlations between centrality measures from a larger and more fluid multi-component network, we have used a method described elsewhere (Boily et al., 1997, 1999; Poulin et al., 1997) to simulate sexual networks of given size N, distribution in sexual activity (DSA, i.e., distribution of degrees) and mixing pattern (MP, who has contact with whom) (Gupta et al., 1989). Moreover, for fixed N–DSA–MP, the simulation method, based on an adaptable computer algorithm controlled by a ''tuning'' parameter, allows the generation of networks with different structural characteristics such as the number of components (NC) and the size of the biggest component (CMax). To limit our last analysis to a manageable computational time given that

Table 9
Pearson correlation matrix between various centrality indices applied to the multi-components network of Fig. 1

	C_{CSize}	C_{Deg}	C_{Ey}	C_{Cs}	C_{Inf}	C'_{Inf}	C_{Bon}	C_{λ}	C_{CN}	C_{CNGr}	C'_{CN}	C''_{CN}
C_{CSize}	1.000	0.411	−0.490	−0.048	0.016	0.322	−0.699	0.554	0.000	0.554	0.463	0.720
C_{Deg}	0.411	1.000	−0.805	0.730	0.797	0.898	0.045	0.836	0.512	0.836	0.982	0.898
C_{Ey}	−0.490	−0.805	1.000	−0.608	−0.637	−0.754	0.074	−0.767	−0.391	−0.767	−0.845	−0.805
C_{Cs}	−0.048	0.730	−0.608	1.000	0.857	0.705	0.204	0.566	0.438	0.566	0.710	0.457
C_{Inf}	0.016	0.797	−0.637	0.857	1.000	0.927	0.070	0.795	0.161	0.795	0.753	0.558
C'_{Inf}	0.322	0.898	−0.754	0.705	0.927	1.000	−0.125	0.946	0.144	0.946	0.871	0.779
C_{Bon}	−0.699	0.045	0.074	0.204	0.070	−0.125	1.000	−0.395	0.653	−0.395	0.024	−0.199
C_{λ}	0.554	0.836	−0.767	0.566	0.795	0.946	−0.395	1.000	−0.000	1.000	0.837	0.836
C_{CN}	0.000	0.512	−0.391	0.438	0.161	0.144	0.653	−0.000	1.000	−0.000	0.545	0.420
C_{CNGr}	0.554	0.836	−0.767	0.566	0.795	0.946	−0.395	1.000	−0.000	1.000	0.837	0.836
C'_{CN}	0.463	0.982	−0.845	0.710	0.753	0.871	0.024	0.837	0.545	0.837	1.000	0.932
C''_{CN}	0.720	0.898	−0.805	0.457	0.558	0.779	−0.199	0.836	0.420	0.836	0.932	1.000

Table 10
Pearson correlation matrix between various centrality indices applied to a multi-component simulated network of size $N = 2250$

	C_{CSize}	C_{Deg}	C_{Ey}	C_{Cs}	C_{Inf}	C'_{Inf}	C_{Bon}	C_{λ}	C_{CN}	C_{CNGr}	C'_{CN}	C''_{CN}
C_{CSize}	1.000	0.288	−0.861	−0.721	−0.563	−0.133	−0.753	0.792	0.000	0.792	0.127	0.347
C_{Deg}	0.288	1.000	−0.379	−0.240	−0.160	0.369	−0.201	0.366	0.521	0.366	0.563	0.594
C_{Ey}	−0.861	−0.379	1.000	0.796	0.711	0.163	0.849	−0.910	−0.015	−0.910	−0.154	−0.302
C_{Cs}	−0.721	−0.240	0.796	1.000	0.951	0.491	0.889	−0.859	0.050	−0.859	−0.095	−0.216
C_{Inf}	−0.563	−0.160	0.711	0.951	1.000	0.585	0.826	−0.764	0.073	−0.764	−0.056	−0.128
C'_{Inf}	−0.133	0.369	0.163	0.491	0.585	1.000	0.389	−0.274	0.331	−0.273	0.273	0.278
C_{Bon}	−0.753	−0.201	0.849	0.889	0.826	0.389	1.000	−0.912	0.215	−0.912	0.055	−0.079
C_{λ}	0.792	0.366	−0.910	−0.859	−0.764	−0.274	−0.912	1.000	0.000	1.000	0.161	0.286
C_{CN}	0.000	0.521	−0.015	0.050	0.073	0.331	0.215	0.000	1.000	0.000	0.984	0.903
C_{CNGr}	0.792	0.366	−0.910	−0.859	−0.764	−0.273	−0.912	1.000	0.000	1.000	0.161	0.287
C'_{CN}	0.127	0.563	−0.154	−0.095	−0.056	0.273	0.055	0.161	0.984	0.161	1.000	0.935
C''_{CN}	0.347	0.594	−0.302	−0.216	−0.128	0.278	−0.079	0.286	0.903	0.287	0.935	1.000

matrix-based centrality indices C_{Bon}, C_{Inf}, C'_{Inf} are computer-intensive, a large multi-component network of size $N = 2250$ with all component sizes less or equal to 250 has

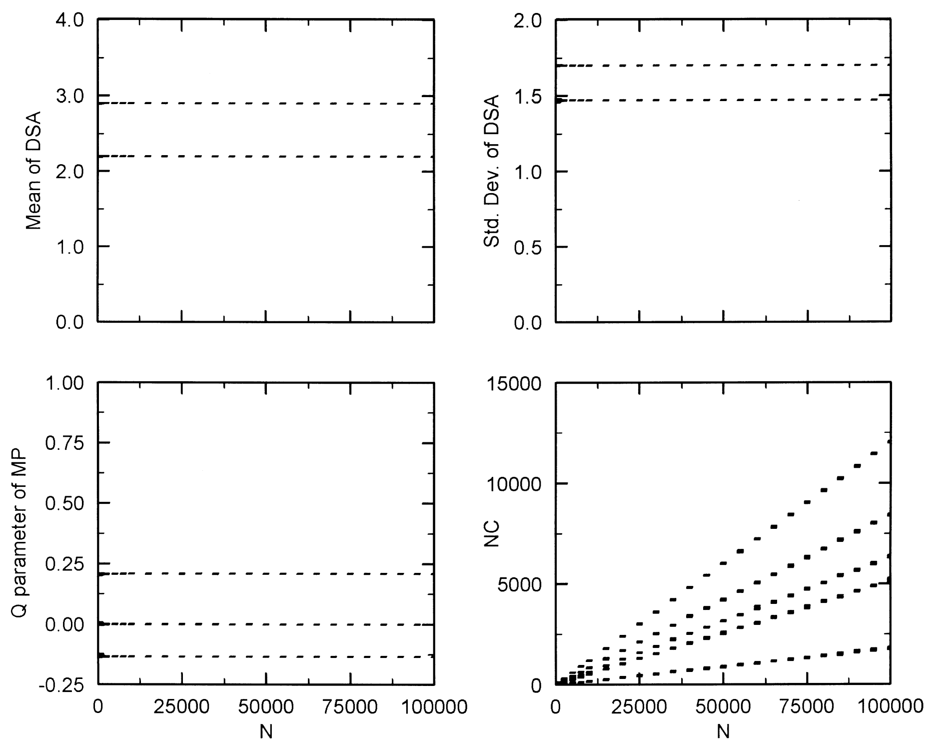

Fig. 4. Relationships between the distribution of sexual activity (DSA), the mixing pattern (MP), the size of the components (NC) and the network size, N, for five types of simulated networks.

been simulated. The centrality indices for points within these networks have been obtained separately for each component before being pooled to obtain the global indices for the correlation analysis of the total network.

Table 10 shows the results of this analysis. For this more heterogeneous network, the range of values for the coefficient of correlation between measures is wider than in the two previous examples. Every centrality index, except ours and C_{Deg}, correlates negatively with the relative size of the components, CSize. This is because the correlation between C_{Deg} and CSize is less important than it was in the network of Fig. 1. The correlation between C'_{CN} and C_{Deg} is now 0.563 compared to 0.982 for the network in Fig. 1. Note that the correlation between $C'_{\mathrm{CN}} - C''_{\mathrm{CN}}$ and C_{CSize} is partly due to the dependence of the nomination growth on C_{CSize}. However, both C'_{CN} and C''_{CN} prove to be sufficiently independent of C_{Deg} and C_{CSize} to be useful as a global measure to characterize networks made of many components of various sizes.

5.2. Graph comparability and scaling law of SMs

In order to study the scaling law associated with the various structural measures introduced in previous sections, we have used the network simulation algorithms mentioned at the end of Section 5.1.

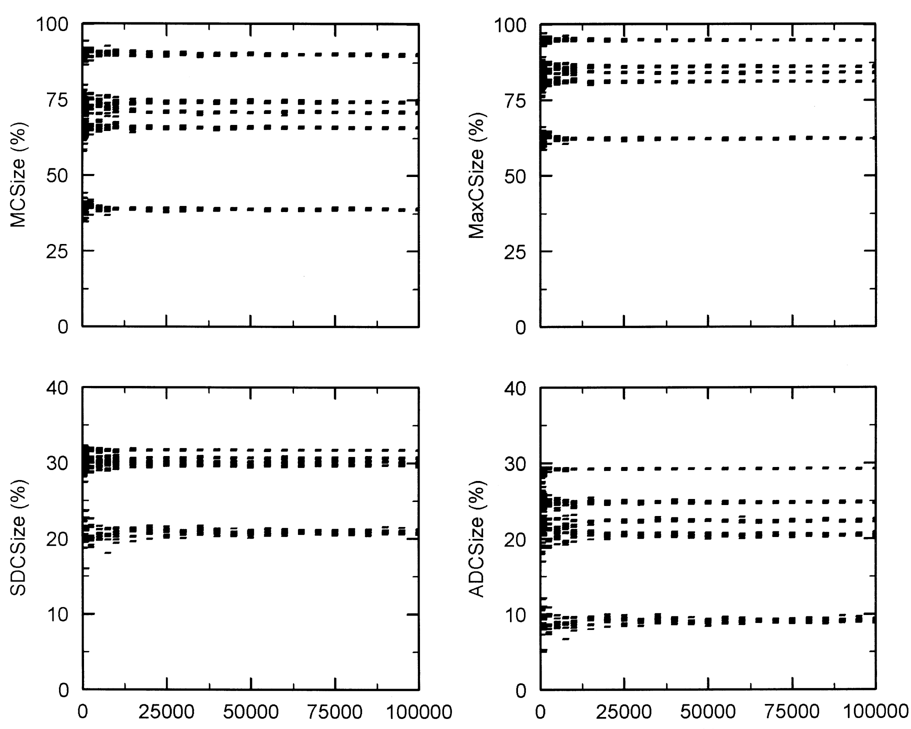

Fig. 5. Same as Fig. 4 for different measures of the components’ size distribution.

Five types of structurally different sexual networks characterized by a selection of combinations among two different DSA and three MP parameters and with sizes varying from $N = 500$ to $N = 100000$ have been simulated. Since the network simulation process is stochastic (the order of partnerships formation is random), 10 repetitions have been performed for each combination of the network characteristics. Finally, various structural measures have been calculated on these simulated networks.

Fig. 4 shows the relationship between the characteristics of the simulated networks (DSA, MP, NC) and N. The DSA, which is the distribution of degrees of the network, is represented by the mean and the standard deviation. Both quantities are constant with N. The same can be said about the MP. MP is summarized by the parameter Q which quantifies the degree of assortativeness of a network (Gupta et al., 1989). Among our five simulated types of networks, three have a proportional mixing pattern with $Q = 0$ (no preference on the choice of partners), one has an assortative mixing with $Q > 0$ (like prefers like) and one has a disassortative mixing pattern with $Q < 0$ (like prefers unlike). Since our networks were generated based on a DSA with four classes of sexual activity (e.g., four different possible degrees), the value of the parameter Q can vary between

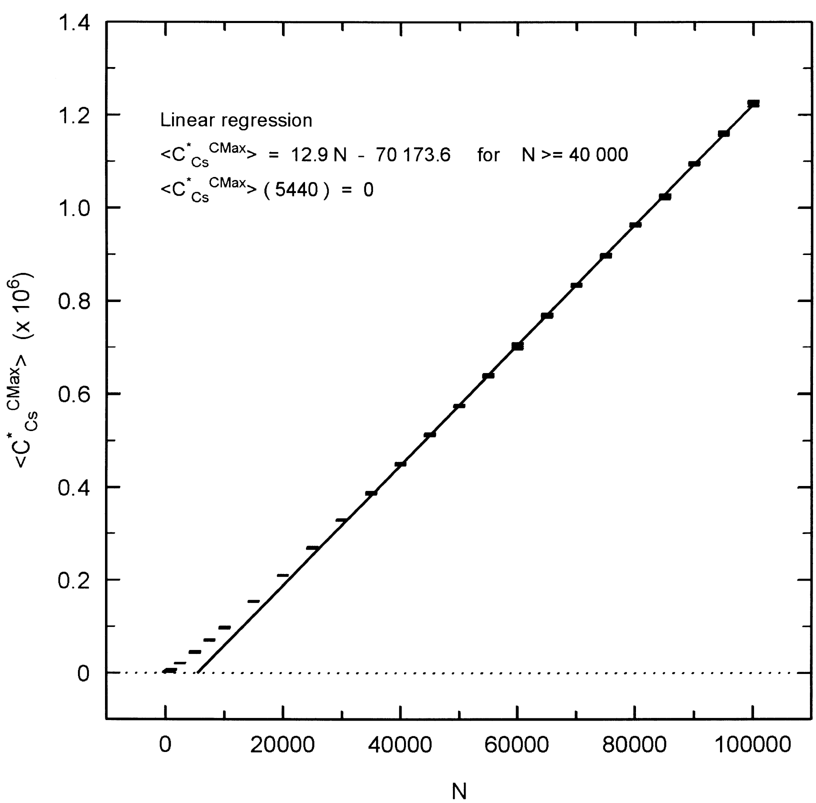

Fig. 6. Scaling law of the closeness measure of centrality for one type of simulated network. For large N, the dependence is linear but not proportional.

$-1/(4-1) = -1/3$ (exclusively disassortative case) and 1 (exclusively assortative case). Fig. 4 also shows that for each type of network, the number of components NC (or CMax; not shown) varies proportionally with the size of the network *N*. These results indicate a good control of the predetermined characteristics of the network during the simulation process and enable the generation of a fixed type of network of various size *N*. Fig. 5 shows that other structural measures characterizing the components' size distribution from an individual point of view (see Section 3.1) are also well controlled during the network-generating process. Those quantities have the advantage of already being scaled with respect to the size of the network *N*.

Measures based on geodesics such as eccentricity (C_{Ey}) and closeness (C_{Cs}) do not vary proportionally in function of the size of our simulated networks (which are assumed to be structurally equivalent for all *N*). Since the calculation of Freeman's centrality indices is very computer-intensive (despite the efficient storing of the network information based on linked list rather than adjacency matrices), these measures have only been calculated for one type of network and five repetitions. Fig. 6 shows the mean distance between individuals ($1/C_{\mathrm{Cs}}$) averaged over all the individuals of the largest component, $\langle C_{\mathrm{Cs}}^{\mathrm{CMax}\,*} \rangle$ as a function of *N*. For large *N*, the functional dependence is

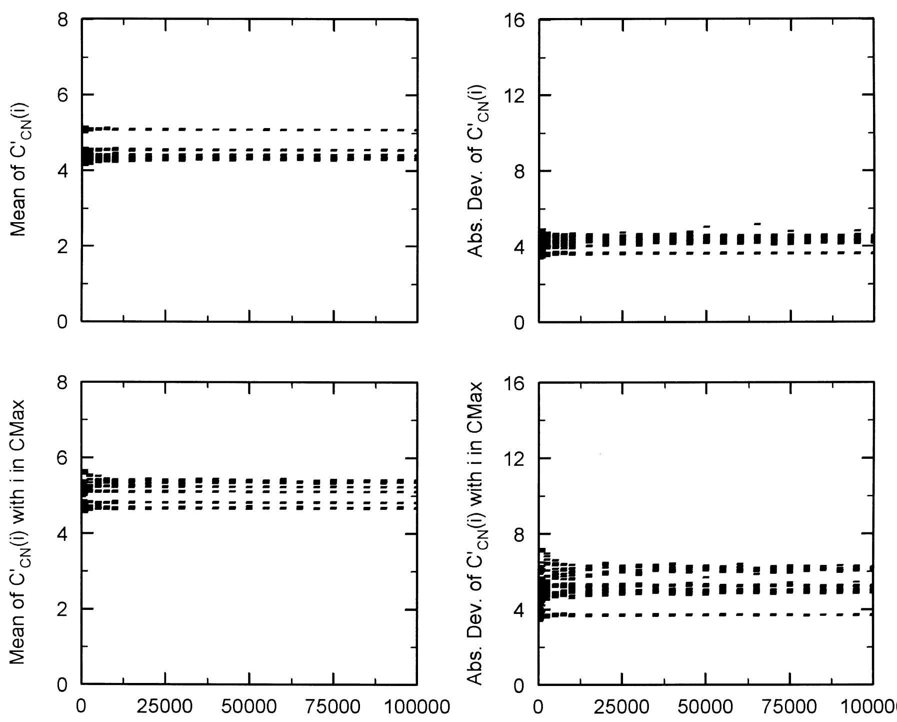

Fig. 7. Dependence on *N* of the multi-components cumulated nominations centrality index, C'_{CN}. Top: statistics are performed on C'_{CN}(i) for the total network. Bottom: statistics are performed on C'_{CN}(i) for the largest component of the network only.

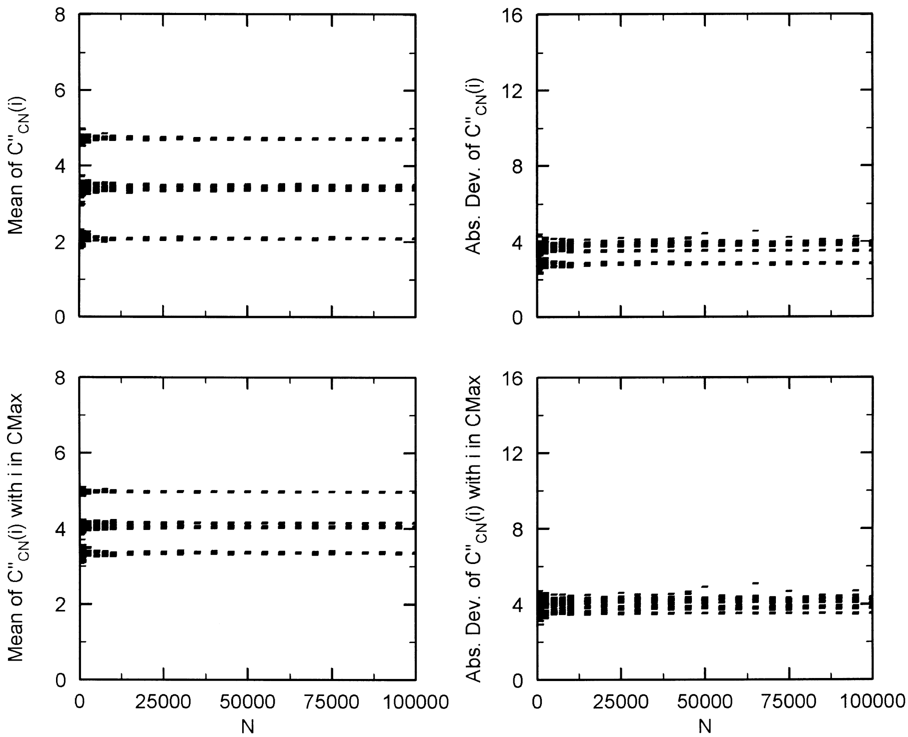

Fig. 8. Same as Fig. 7 for the corrected multi-components cumulated nominations centrality index, C''_{CN}.

linear but not proportional, i.e., the regression line does not go through the origin. To be able to consider the scaling between N and $\langle C_{\mathrm{Cs}}^{\mathrm{CMax}\,*} \rangle$ proportional, the size of networks should be much larger than 5440.

Matrix-based measures (i.e., C_{Bon}, C_{Inf} and C'_{Inf}) have not been considered since, depending on computer resources, they cannot be obtained for networks larger than a few thousands points.

Finally, Figs. 7 and 8 show statistics on C'_{CN} and C''_{CN}, respectively. These measures were studied for the five types of network simulated but only two (with the same DSA and MP) are shown for convenience. Similar results were obtained for the other types. As can be seen, both measures are well suited to describe multi-component networks since they do not depend on N. Moreover, even for large networks, the nomination mapping converges rapidly to the centrality indices (e.g., few minutes on a typical UNIX workstation for a network of size $N = 100{,}000$ with the maximum of $C_{\mathrm{CSize}} \sim 80\%$).

6. Conclusion

In this paper, new computationally efficient centrality measures have been presented. The method, which is based on a cumulated nomination scheme, uses a mapping that explores all the possible paths between any pair of individuals of the network.

An index of centrality is defined to characterize individuals within a connected network and is shown to be related to the Bonacich index of centrality. However, the new index, based on a new mapping, is shown to have a better numerical stability than its Bonacich counterpart. Moreover, the new mapping further permits to define new centrality indices that allow the comparison of centrality scores of individuals in different networks (or components) and, by extension, permit the characterization of multi-component networks.

Our new measures and the more classical ones have been compared when applied to a connected sexual network of 40 homosexual men with AIDS. Our new index proves to have an excellent discriminant power and produces results that are directly interpretable. These results have been verified by a correlation analysis between various centrality indices calculated on both an illustrative and a simulated multi-component network. Our method presents the advantage of not being computer-time-intensive and of limiting the quantity of memory required to a factor proportional to the size of the network. It is therefore suitable to characterize large social networks such as those found in real applications.

The scaling laws of various structural measures have also been examined based on simulated networks. Structural measures based on the distribution of the degrees and of the component sizes are very appropriate to describe a network independently of its size. Classical indices of centrality such as eccentricity and closeness do not obey any simple scaling law in function of the network size. On the contrary, our new indices of centrality do not show any functional dependence on the size of the network. Henceforth, they are suitable to characterize networks made of many components with the advantage, over those based on the components' size, that the components structure is being considered globally.

As a next step, it would be very interesting and relevant to examine and compare the discriminant power of our measure (based on real data or simulation data) to identify individuals at higher risk of STD or HIV infections in sexual or needle sharing networks. Although not completely equivalent, the cumulated nominations scheme explores a social network and nominates individuals in a similar fashion than viral or bacterial infection may circulate through the network of contacts and potentially infect individuals.

Acknowledgements

M.-C.B. and B.M. acknowledge the financial support of NHRDP and FRSQ.

References

Altmann, M., 1993. Reinterpreting network measures for models of disease transmission. Soc. Networks 15, 1–17.

Altmann, M., Wee, B.C., Willard, K., Peterson, D., Gatewood, L.C., 1994. Network analytic methods for epidemiological risk assessment. Stat. Med. 13, 53–60.

Anderson, R., 1988. The use of mathematical models in the epidemiological study of infectious diseases and in the design of mass immunization programs. Epidemiol. Infect. 101, 1–10.

Anderson, R., 1991. The transmission dynamics of sexually transmitted diseases: the behavioral component. In: Wasserheit, J. (Ed.), Research Issues in Human and Sexually Transmitted Diseases in the AIDS Era. American Society for Microbiology, Washington, pp. 38–60.

Anderson, R., Medley, G., May, R., Johnson, A. et al., 1986. A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS. IMA J. Math. Appl. Med. Biol. 3, 229–263.

Anderson, R., Blythe, S., Gupta, S., Koning, E., 1989a. The transmission dynamics of the human immunodeficiency virus type 1 in the male homosexual community in the United Kingdom: the influence of changes in sexual behaviour. Philos. Trans. R. Soc. London, Ser. B 325, 45–58.

Anderson, R., Ng, T., Boily, M.-C., May, R., 1989b. The influence of different sexual contact patterns between age classes on the predicted demographic impact of AIDS in developing countries. N. Y. Acad. Sci. 569, 240–274.

Anderson, R.M., Gupta, S., Ng, W.T., 1990. The significance of sexual partner choice networks for the transmission dynamics of HIV. J. AIDS 3, 417–429.

Anderson, R., May, R., Boily, M.-C., Garnett, G., 1991. The spread of HIV-1 in Africa: sexual contact patterns and the predicted demographic impacts of AIDS. Nature 352, 581–589.

Auerbach, D., Darrow, W., Jaffe, H., Curran, J., 1984. Cluster of cases of the acquired immune deficiency syndrome: patients linked by sexual contact. Am. J. Med. 76, 487–492.

Boily, M.-C., Anderson, R., 1991. Sexual contact patterns between men and women and the spread of HIV-1 in urban centers in Africa. IMA J. Math Appl. Med. Biol. 8, 221–247.

Boily, M.-C., Mâsse, B., Poulin, R., Alary, M., 1997. Estimation of the mixing pattern from large networks. In: The 12th Meeting of the International Society of Sexually Transmitted Diseases Research (ISSTDR), Seville, Spain. Abstract O 124.

Boily, M.-C., Poulin, R., Mâsse, B., Alary, M., 1999. The estimation of the mixing matrix from large sexual networks. 1999. Submitted.

Bonacich, P., 1972. Factoring and weighting approaches to status scores and clique identification. J. Math. Sociol. 2, 113–120.

Bonacich, P., 1987. Power and centrality: a family of measures. Am. J. Soc. 92, 1170–1182.

Devaney, R.L., 1989. An Introduction to Chaotic Dynamical Systems. Addison-Wesley, Don Mills, ON.

Dietz, K., 1988. On the transmission dynamics of HIV. Math. Biosci. 90, 397–414.

Donninger, C., 1986. The distribution of centrality in social networks. Soc. Networks 8, 191–203.

Doyle, P., Snell, J., 1984. Random walks and electrical networks. In: The Carus Mathematical Monographs 22 The Mathematical Association of America, Washington, DC.

Freeman, L.C., 1979. Centrality in social networks: I. Conceptual clarification. Soc. Networks 1, 215–239.

Friedman, S.R., Neaigus, A., Jose, B., Curtis, R., Goldstein, M., Ildefonso, G., Rothenberg, R.B., Des Jarlais, D.C., 1997. Sociometric risk networks and risk for HIV infection. Am. J. Public Health 87, 1289–1296.

Gupta, S., Anderson, R.M., May, R.M., 1989. Networks of sexual contacts : implications for the pattern of spread of HIV. AIDS 3, 807–817.

Harary, F., 1969. Graph Theory. Addison-Wesley, Reading, MA.

Hethcote, H., Yorke, J., 1984. Gonorrhea transmission dynamics and control. In: Levin, S. (Ed.), Lecture Notes in Biomathematics 56 Springer-Verlag, New York, p. 99.

Iannelli, M., Milner, F., Pugliese, A., Gonzo, M., 1997. The HIV/AIDS epidemics among drug injectors: a study of contact structure through a mathematical model. Math. Biosci. 139, 25–58.

Kleinbaum, D.G., Kupper, L.L., Morgenstern, H., 1982. Epidemiologic Research: Principles and Quantitative Methods. Van Nostrand-Reinhold, New York.

Klovdahl, A.S., 1985. Social networks and the spread of infectious diseases: the AIDS example. Soc. Sci. Med. 21, 1203–1216.

Klovdahl, A.S., Potterat, J.J., Woodhouse, D.E., Muth, S.Q., Darrow, W.W., 1994. Social networks and infectious diseases: the Colorado Springs study. Soc. Sci. Med. 38, 79–88.

Kretzschmar, M., Wiessing, L., 1998. Modelling the spread of HIV in social networks of IVDU. AIDS 12, 801–811.

Kretzschmar, M., Jager, J.C., Reinking, D.P., van Zessen, G., Brouwers, H., 1995. A network modeling approach for assessing the efficiency of prevention strategies. J. Math. Sociol. 20, 351–374.

Poulin, R., Boily, M.-C., Mâsse, B., Alary, M., 1997. Large sexual networks and their structural measures. In: The 12th Meeting of the International Society of Sexually Transmitted Diseases Research (ISSTDR), Seville, Spain. Abstract O 125.

Rothenberg, R.B., Potterat, J.J., Woodhouse, D.E., Darrow, W.W., Muth, S.Q., Klovdahl, A.S., 1995. Choosing a centrality measure: epidemiologic correlates in the Colorado Springs study of social networks. Soc. Networks 17, 273–297.

Rothenberg, R.B., Sterk, C., Toomey, K.E., Potterat, J.B., Johnson, D., Schrader, M., Hatch, S., 1998. Using social network and ethnographic tools to evaluate syphilis transmission. Sex. Transm. Dis. 25, 154–160.

Stephenson, K., Zelen, M., 1989. Rethinking centrality: methods and examples. Soc. Networks 11, 1–37.

Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge Univ. Press, Cambridge, No. 8 in Structural Analysis in the Social Sciences.