# Community structure in social and biological networks

**M. Girvan*[†‡] and M. E. J. Newman*[§]**

* , 13 , 501 , 14 53-2501 4 10 -1120

6, 2002 ( 6, 2001)

**A number of recent studies have focused on the statistical properties of networked systems such as social networks and the Worldwide Web. Researchers have concentrated particularly on a few properties that seem to be common to many networks: the small-world property, power-law degree distributions, and network transitivity. In this article, we highlight another property that is found in many networks, the property of community structure, in which network nodes are joined together in tightly knit groups, between which there are only looser connections. We propose a method for detecting such communities, built around the idea of using centrality indices to find community boundaries. We test our method on computer-generated and real-world graphs whose community structure is already known and find that the method detects this known structure with high sensitivity and reliability. We also apply the method to two networks whose community structure is not well known—a collaboration network and a food web—and find that it detects significant and informative community divisions in both cases.**

Many systems take the form of networks, sets of nodes or vertices joined together in pairs by links or edges (1). Examples include social networks (2 4) such as acquaintance networks (5) and collaboration networks (6), technological networks such as the Internet (7), the Worldwide Web (8, 9), and power grids (4, 5), and biological networks such as neural networks (4), food webs (10), and metabolic networks (11, 12). Recent research on networks among mathematicians and physincreasingly large components (connected subsets of vertices), which are taken to be the communities. Because the components are properly nested, they all can be represented by using a tree of the type shown in Fig. 2, in which the lowest level at which two vertices are connected represents the strength of the edge that resulted in their first becoming members of the same community. A "slice" through this tree at any level gives the communities that existed just before an edge of the corresponding weight llt6.3(a-30i)-49.9(isted)-(t)-30(ypec)-33r0(ypec)-3sometim01(ll)-Tentedising

1. Calculate the betweenness for all edges in the network.

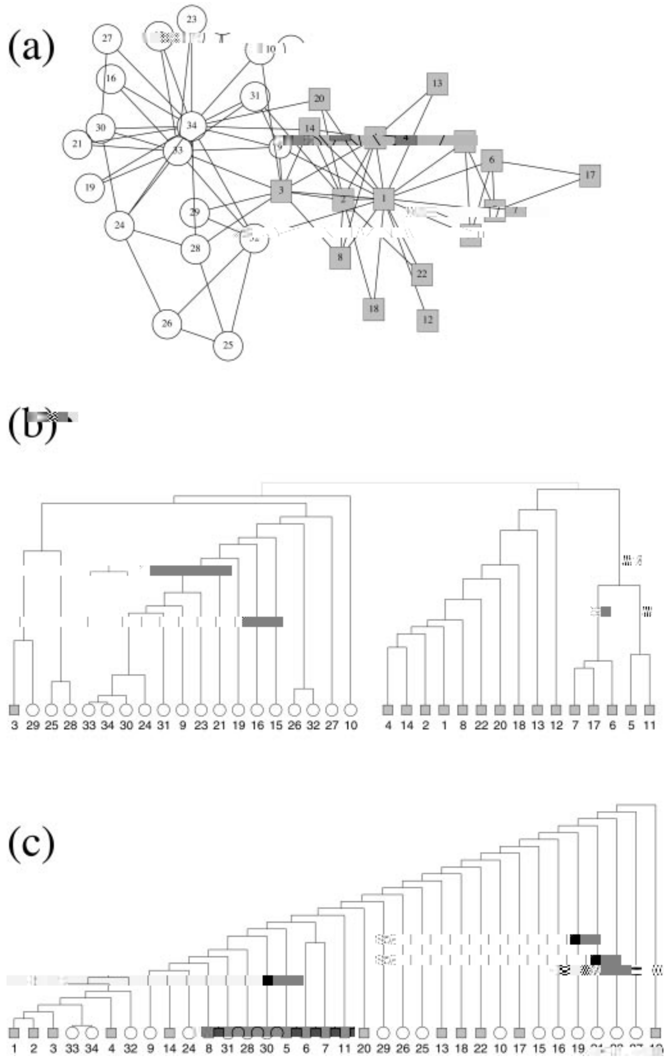


**Fig. 4.** (*a*) … (26) … (*b*) … (*c*) …

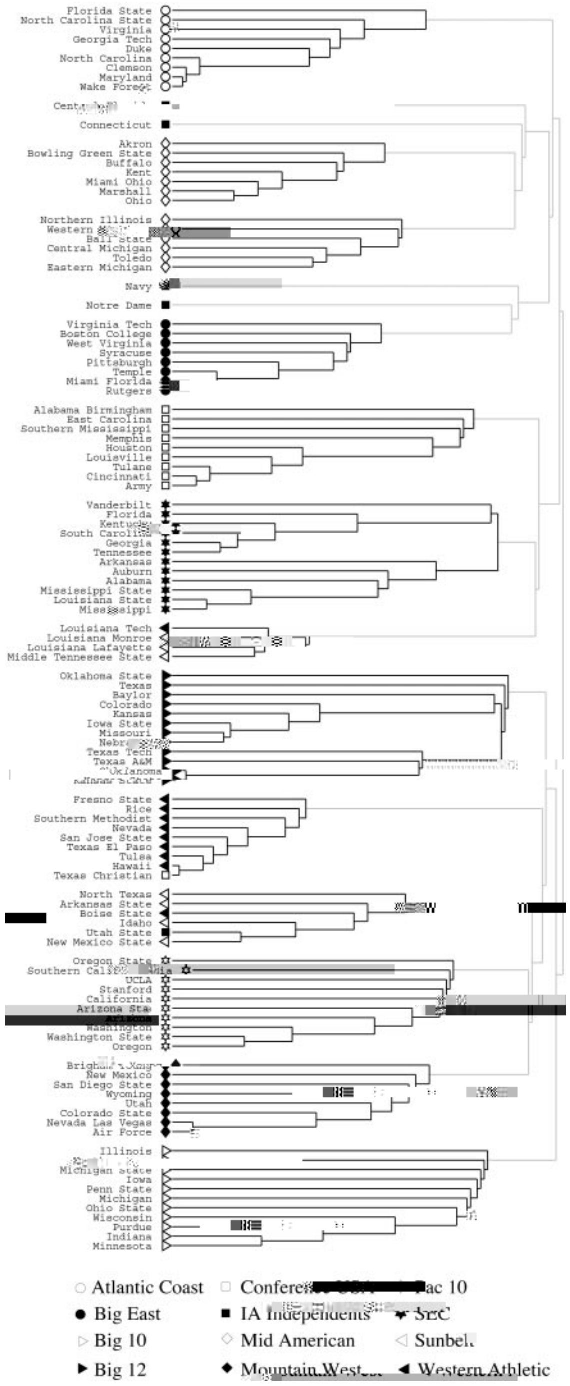


**Fig. 5.**

interesting is that it incorporates a known community structure. The teams are divided into conferences containing around 8 12 teams each. Games are more frequent between members of the same conference than between members of different conferences, with teams playing an average of about seven intraconference games and four interconference games in the 2000 season. Interconference play is not uniformly distributed; teams that are geographically close to one another but belong to different conferences are more likely to play one another than teams separated by large geographic distances.

Applying our algorithm to this network, we find that it identifies the conference structure with a high degree of success (Fig. 5). Almost all teams are correctly grouped with the other teams in their conference. There are a few independent teams that do not belong to any conference these tend to be grouped with the conference with which they are most closely associated. The few cases in which the algorithm seems to fail actually correspond to nuances in the scheduling of games. For example, the Sunbelt Conference is broken into two pieces and grouped with members of the Western Athletic Conference. This happens because the Sunbelt teams played nearly as many games against Western Athletic teams as they did against teams in their own conference. They also played quite a large fraction of their interconference games against Mid-American teams. Naturally, our algorithm fails in cases like this where the network structure genuinely does not correspond to the conference structure. In all other respects, however, it performs remarkably well.

## A

In the previous section we tested our algorithm on a number of networks for which the community structure was known beforehand. The results indicate that our algorithm is a sensitive and accurate method for extracting community structure from both real and artificial networks. In this section, we apply our method to two more networks for which the structure is not known and show that in these cases it can help us to understand the make-up of otherwise complex and tangled datasets. Our first example is a collaboration network of scientists; our second is a food web of marine organisms in the Chesapeake Bay.

**C , Ne , .** We have applied our community-finding method to a collaboration network of scientists at the Santa Fe Institute, an interdisciplinary research center in Santa Fe, New Mexico (and current academic home to both authors of this article). The 271 vertices in this network represent scientists in residence at the Santa Fe Institute during any part of calendar year 1999 or 2000 and their collaborators. An edge is drawn between a pair of scientists if they coauthored one or more articles during the same time period. The network includes all journal and book publications by the scientists involved, along with all papers that appeared in the institute's technical reports series. On average, each scientist coauthored articles with approximately five others.

In Fig. 6 we illustrate the results from the application of our algorithm to the largest component of the collaboration graph (which consists of 118 scientists). Vertices are drawn as different shapes according to the primary divisions detected. We find that the algorithm splits the network into a few strong communities, with the divisions running principally along disciplinary lines. The community indicated by diamonds is the least well defined and represents a group of scientists using agent-based models to study problems in economics and traffic flow. The algorithm further divides this group into smaller components that correspond roughly with the split between economics and traffic. The community represented by circles is comprised of a group of scientists working on mathematical models in ecology and forms a fairly cohesive structure, as evidenced by the fact that the algorithm does not break it into smaller components to any significant extent. The largest community, indicated by squares of various shades, represents a group working primarily in statist-224.8cferphysnentsc7(physnents)-30s19.9(serphysnent25.4f.5(ev)sed)4f.5
with the 263 TDhe263 Taga39.8(263 Tarosents(263 Tration263 Tinte-576.7(g)9

ifhdroph36.3133 (-5larf-2shipsT320. F3 it)f2 (963 0 )-30 Tct( )Tj. F13 it
with,ne266h8TpluithwTJ2to4.5.belohemati4hms,
withdoefooewebs(28),wh36hteneto0(701d2(3lust)r)-901d2(tif)-30xamw901d2(a

in mind, and it is possible that it may not perform as well on dense networks.

## C

In this article we have investigated community structure in networks of various kinds, introducing a method for detecting such structure. Unlike previous methods that focus on finding