Ernst Moritz Arndt Universität Greifswald 12. 6. 2009

Komplexe Netzwerke

Netzwerkstrukturen

Dr. Matthias Scholz

www.network-science.org/SS2009.html

6 Netzwerkstrukturen

Soziologie: Gesellschaftsstruktur (*community structure*)

Gibt es Bereiche in einem Netzwerk, die stärker vernetzt sind als andere? Existieren stark vernetzte Cluster oder Gruppen? Modularität? (Newman, 2006)

6.1 Dichte

Die Dichte der Vernetzung beschreibt das Verhältnis der vorhandenen Beziehungen zur Zahl der maximal möglichen Beziehungen in einem Netzwerk. Die Dichte kann sich lokal auf einzelne Knoten und deren Nachbarschaft beziehen oder global auf das gesamte Netzwerk.

Kantendichte (*edge density*) ist das Verhältnis der Anzahl m von Kanten im Graphen G zur Gesamtzahl $\frac{1}{2}N(N-1)$ der möglichen Kanten (bei N Knoten).

$$D = \frac{m}{\frac{1}{2}N(N-1)}$$

Dreiecksdichte (*triangle density*) ist das Verhältnis der Anzahl aller Dreiecke im Graphen G zur Gesamtzahl der möglichen Dreiecke bei N Knoten. Als Dreieck werden drei Knoten bezeichnet, die vollständig miteinander vernetzt sind.

$$D = \frac{Anzahl Dreiecke}{\frac{1}{6}N(N-1)(N-2)}$$

6.1.1 Cluster-Koeffizient

Mit dem *Cluster-Koeffizienten* (Watts and Strogatz, 1998) kann man die lokale Kompaktheit oder Dichte eines Graphen bestimmen.
Er beschreibt, wie stark Nachbarknoten auch untereinander vernetzt sind.

Lokal: Der Cluster-Koeffizient C_i bezieht sich auf einen einzelnen Knoten v_i.
C_i entspricht der Wahrscheinlichkeit eines Knoten v_i, dass zwei seiner Nachbarknoten

selbst auch miteinander verbunden sind.
Sind Freunde von mir auch untereinander befreundet? $\longrightarrow$ 6.1.2 Transitivität.

Der Cluster-Koeffizient ist gegeben durch die Anzahl der vorhandenen Kanten zwischen Nachbarn vom Knoten v_i im Verhältnis zur Anzahl möglicher Kanten.

$$C_i = \frac{\sum_k \sum_j a_{ij} a_{ik} a_{jk}}{k_i(k_i - 1)}$$

Global: Wie stark ist die Struktur eines gesamten Graphen in Cluster aufgeteilt?
Dazu kann man den gemittelten (globalen) Cluster-Koeffizienten des gesamten Graphen berechnen:

$$C = \frac{1}{N} \sum_{i}^{N} C_i$$

Vollständig vernetzte Bereiche im Graphen ($C_i = 1$) werden als *Clique* bezeichnet.

Clique: Eine *Clique* ist ein vollständiger Teilgraph. Jeder Knoten ist mit jedem anderen Knoten verbunden. Eine Clique mit N Knoten besitzt alle $\frac{N(N-1)}{2}$ möglichen Kanten.

6.1.2 Transitivität

Eine transitive Beziehung ist gegeben wenn aus A $\longrightarrow$ B und B $\longrightarrow$ C stets A $\longrightarrow$ C folgt (Schnittregel).

6.2 Bipartite Graphen

Oft hat man zwei (oder sogar mehr) verschiedene Arten von Netzwerk-Elementen, die man gemeinsam betrachten möchte.
Beispielsweise:

- *Gene* und *Funktionen oder Krankheiten*, an denen verschiedene Gene beteiligt sind;
- *Wissenschaftler und gemeinsame Projekte/Publikationen* oder
- *Schauspieler* und *Filme.*

Als Modell für Netzwerke mit zwei verschiedenen Arten von Elementen können bipartite Graphen benutzt werden. Bipartite Graphen haben zwei verschiedene Mengen von Knoten. Die Kanten gehen nur von Knoten der einen Menge (*Gene*) zu Knoten der andernen Menge (*Funktionen*).

bipartit: Ein Graph heißt *bipartit*, wenn sich seine Knoten so in zwei Mengen aufteilen lassen, dass es nur Kanten zwischen den Knoten der einen Menge und denen der anderen Menge gibt.

Bipartite Graphen lassen sich auch in zwei 'normale' *unipartite* Graphen mit jeweils nur einer Menge von Knoten überführen.

Beispiel Gen-Gen-Netzwerk: zwei Gene sind mit einer Kante verbunden, wenn sie an der selben biologischen Funktion beteiligt sind. Die gemeinsamen Funktionen werden dabei nicht als Knoten dargestellt, bestimmen aber indirekt die Kanten und damit die Clusterstruktur im Netzwerk.

Literatur

Newman, M.E.J. Modularity and community structure in networks. *Proc. Natl. Acad. Sci.*, 103(23):8577–8582, 2006.

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