Forschungsmodul: Komplexe Systeme

Bericht zur Vorlesung vom 06. Dezember 2007 von Jan-Philip Gehrcke

Charakterisierung komplexer Netze

Die Eigenschaften komplexer Netze sind vielfältig. Selbst ein auf die Knoten-und-Kanten-Topologie beschränktes Netzwerk kann viele besondere Charakteristiken aufweisen. Bestimmte Eigenschaften, die bei Visualisierung eines Netzes sofort auffallen würden, sind auch mathematisch berechenbar. Hier sei beispielsweise der Assortativitätskoeffizient genannt, welcher besprochen wird. Auch sollen Cluster-Eigenschaften, das Barabasi-Albert-Modell und die Epidemieanfälligkeit eines Netztes diskutiert werden.

Assortativität

Er beschreibt anschaulich, ob in einem Netzwerk eher gleichartige Knoten (bezogen auf den Kantengrad) miteinander vernetzt sind oder ob sich eine Mischung ausbildet, in der Knoten mit vielen Kanten eher zu schlecht vernetzten Knoten verbunden sind. Dieses Problem ist einfach zu berechnen. Ein positiver Koeffizient bedeutet eine klare Struktur, in der vielzahlig bestückte Knoten vorherrschen. In einem Netz negativer Assortativität hingegen sind keine so deutlichen “Master-Knoten” vorhanden; es wirkt durchmischter. Ganz allgemein ist der Assortativitätskoeffizient jedoch auf jede Art von Auslegung des Begriffs der “Gleichartigkeit” anwendbar. So ist er auch für das Heiratsverhalten der Menschen berechenbar: Sind eher weiße mit weißen verheiratet oder schwarze mit schwarzen oder gewisse Mischungen vorhanden? Der Assortativitätskoeffizient kann das auf der Grundlage von einer Menge an Daten aussagen.

Für soziale und ähnliche Netzwerke liegen fast grundsätzlich positive Koeffizienten vor. Eine stark frequentierte Person zieht mehr andere gut vernetzte Personen an, als Menschen, die eigentlich gar keine echten Knoten sind, sondern alleine stehen im Netzwerk. Betrachtet man hingegen biologische oder technische Systeme, stellt man überwiegend negative Assortativitäten fest. Anschaulich gesagt ist hier “deutlicher” alles mit allem verknüpft und es bilden sich kaum “Machtzentren”, welche doch eher menschlich sind, wie auch eine “Klassentrennung”.

Clustering

Der Clusteringkoeffizient c eines Netzes beschreibt die Konnektivität der Nachbarn untereinander. Für ein “mikrokanonisches” Netz kann man ihn exakt berechnen, im kanonischen System entspricht er einleuchtender Weise der Wahrscheinlichkeit, dass überhaupt Kopplungen existieren. In sozialen Beziehungsnetzen sollte er relativ hoch sein, da die Bekannten einer Person untereinander normalerweise auch bekannt sind. Netzwerke, die die “Small World” - u.a. ein Modell für das Beziehungssystem der Menschen auf der Erde - beschreiben sollen, brauchen also einen relativ hohen Clusteringkoeffizienten.

Clustering in “Small World”-Netzen:

Einen sehr hohen Clusteringkoeffizienten in einem überschaubaren Netzwerk kann man erzeugen, indem man die Knoten kreisförmig anordnet und jeden in einer definierten Vorwärtsrichtung (links- oder rechtsherum) mit seinen k nächsten Nachbarn verbindet. Dann gilt:

$$c = \frac{3(k-1)}{2(2k-1)}$$

Die zweite Anforderung eines Small World Netzes ist - wie bereits angesprochen wurde - eine sehr kurze mittlere Weglänge l zwischen beliebigen Knoten. Man erinnere sich an die angeblich nur sechs benötigten Beziehungs-Hops innerhalb der Menschheit, um zu einem beliebigen Menschen zu gelangen. Im bisher beschriebenen Netz ist l groß (“Large World”). Numerisch wurde gezeigt:

$$l = \frac{N}{4k}$$

Nimmt man aber nun an, dass beliebige existierende Verbindungen mit der Wahrscheinlichkeit p getrennt und stattdessen zwischen zufälligen Knoten wieder aufgebaut werden können, erreicht man schon für kleine p viel kleinere l. Denn anschaulich werden nun Wege möglich, die "quer" durch das Kreisnetzwerk ziehen. l skaliert dann allgemein wie $l \propto \ln(N)$. Eine exakte Approximation liefert:

$$l = \frac{N}{2k * \sqrt{(Nkp)^2 + 2Nkp}} arctanh(\frac{Nkp}{Nkp+2})$$

Der Clusteringkoeffizient nimmt dabei nur mit $(1-p)^3$ ab, sodass die Eigenschaft eines hohen Clusteranteils für kleine p erhalten bleibt. Es existiert im beschriebenen Kreismodell also ein (kleines) p, für das l und c die "Small World"-Eigenschaften gleichzeitig erfüllen.

Barabasi-Albert Modell

Das erst 1999 entwickelte Modell soll die Entstehung - also das Wachstum - eines Netzes beschreiben, welches einem Potenzgesetz folgt. Letztmalig wurde bereits diskutiert, dass die Kantenzahl pro Knoten beispielsweise für das Internet potenz- statt poissonverteilt ist.

In jedem Takt- oder Zeitschritt t soll ein Knoten mit m Verbindungen zu dem Netzwerk dazustoßen. Wo er anknüpft, ist jedoch nicht ganz dem Zufall überlassen - er soll sich lieber mit Knoten verbinden, die schon viele Verbindungen haben. Dies entspricht in gewisser Weise schon dem Internet - populäre Seiten werden eher verlinkt als andere. Die Wahrscheinlichkeit $\Pi(k_i)$, dass ein neues Element sich an den Knoten i mit k_i Kanten anfügt, sei also gegeben durch

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \quad .$$

Betrachtet man dieses Modell im Kontinuumslimes, kann man eine Differentialgleichung für das Kantenwachstum des Knotens i aufstellen. Mit der Anfangsbedingung, dass ein Knoten zur Zeit t_i zum Netzwerk dazugestoßen ist und zu Beginn m Kanten hatte, lautet die Lösung:

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{1/2}$$

Rechnet man unter dieser Vorraussetzung weiter und macht eine Wahrscheinlichkeitsbetrachtung der Anzahl der Kanten k pro Knoten - also der Knotengradverteilung - erhält man das gewünschte Potenzgesetz:

$$P(k) \propto \frac{f(t)}{k^3}$$

f ist zeitabhängig und enhält die Konstante m.

SIR Modell: Susceptible / Infected / Recovered (Removed)

Aufschlussreich ist eine Untersuchung eines komplexen Netzwerks in Hinsicht der Anfälligkeit für Viren. Dabei sei der Begriff des Virus' allgemein zu verstehen. Klassifizieren kann man die Knoten dann in drei Gruppen, nämlich:

- gesund aber anfällig (susceptible): Menge S
- angesteckt (infected): Menge I
- geheilt (recovered) oder aus dem Netz entfernt (removed): Menge R

Mit einer Ansteckungsrate β und einer Heilungs- oder Sterberate γ (was in dieser Abstraktion keinen Unterschied macht bis auf removed/recovered) kann man dann Differentialgleichungen für die Dynamik des Gesamtsystems aufstellen:

$$\dot{S} = -\beta I S \quad ; \quad \dot{I} = \beta I S - \gamma I \quad ; \quad \dot{R} = \gamma I$$

Mithilfe dieser Gleichungen kann man bestimmte Szenarien simulieren und z.B. feststellen, ob ein Netzwerk die “Viruswelle” übersteht (nach endlicher Zeit Menge I wieder leer) oder einer Epidemie zum Opfer fällt (alle Knoten in I). Rein rechnerisch ergibt sich Folgendes: Eine Infektion in einem Netz mit mittlerem Überschussknotengrad $\langle d \rangle \geq 1$ führt zwangsläufig zu einer Epidemie. Dies ist damit begründet, dass ein Angesteckter im Mittel mindestens einen Kontakt hat, der noch nicht angesteckt ist. Somit kann sich der Virus ungehindert im gesamten Netz ausbreiten.

Derartige Betrachtungen helfen auch optimale Wege zum Schutz vor einer möglichen Epidemie - durch zum Beispiel Impfungen - zu finden. Basierend auf dem Modell eines komplexes Netzwerks lassen sich geschickte Impfstrategien entwickeln, die eine Epidemie bei minimalem Impfaufwand verhindern können.