„Maximizing the Spread of Influence through a social network“
von Jan Koopmann

Datamining Seminar SS06

Auf Basis von
„Maximizing the Spread of Influence through a social network“
von David Kempe, Jon Kleinberg, Éva Tardos

Worum geht es?

Viele Bereiche in denen Mundpropaganda eine Rolle spielt:

- Verbreitung eines neuen Produkts(Handy, Getränk, Notebook) unter Jugendlichen.
- Die Einführung einer neuen Arznei bei Ärzten.
- Politische Bewegungen

=>Produkt kann sich weiterverbreiten, kann aber auch stagnieren!

Viral Marketing

- Setzt auf Mundpropaganda:
- Kleine einflussreiche Startgruppe wird mit Produkt bekannt gemacht
- Diese machen Freunde mit dem Produkt bekannt,
- Welche wieder Freundesfreunde „infizieren“

Wie wählt man die Startergruppe aus?

Gliederung:

1. Basisannahmen und Modelle
2. Komplexität und Approximierbarkeit
3. Allgemeine Modelle
4. Nichtprogressive Prozesse
5. Allgemeine Marketing Strategien

1.Basisannahmen

Soziales Netzwerk:

- Gerichteter Graph G
- Knoten sind entweder aktiv oder inaktiv
- Tendenz aktiviert zu werden nimmt zu je mehr Nachbarn aktiviert sind
- aktiv → intaktiv
- Initiale Menge von aktiven Knoten A_0

1.1 Linear Treshold-Model

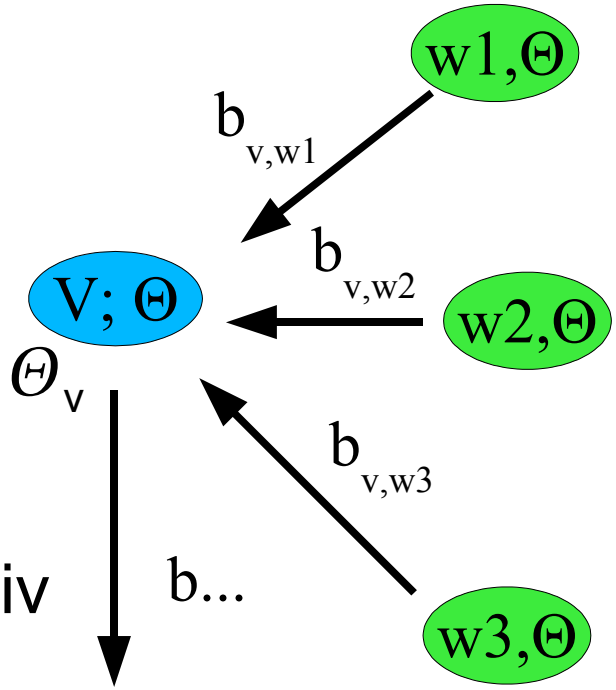
Seien w in Schritt t-1 aktive Nachbarn von Knoten v

$$\sum_{w \text{ Nachbar von } v} b_{v,w} \leq 1$$

v in Schritt t aktiv, wenn $\sum_{w \text{ Nachbar von } v} b_{v,w} \geq \Theta_v$

Knoten die in $t-1$ aktiv sind auch in t aktiv

$\Theta_v \in [0,1]$ und zufällig gewählt für jeden Knoten



1.2 Independent Cascade-Modell

- diskrete Schritte
- einmaliger Versuch Nachbarn zu aktivieren
- $p_{v,w}$ WK, dass Knoten v Nachbarknoten w aktiviert

1.3 Einflussfunktion

σ(A):=Erwartete Anzahl an aktivierten Knoten am Ende des Prozesses = Einfluss einer Knotenmenge

Influence Maximization Problem:= Es ist eine k-elementige Menge an Knoten zu finden für die σ(A) maximiert wird.

2. Komplexität und Approximierbarkeit

Beide Algorithmen NP-schwer.
Gute Approximation mittels *Greedy Hill-Climbing*(Faktor 1-1/e≈63%)

Greedy Hill-Climbing:=Setze Bergsteiger irgendwo in Gebirge ab. Dieser wandert bergauf zum nächsten Gipfel und stoppt dort.
=>lokales Maximum

2.1. Approximierbarkeit

Sei $f: \hat{U} \rightarrow x \in \mathbb{R}, \hat{U} \subseteq U$

Def: f` ist submodular, wenn *diminishing returns* Eigenschaft erfüllt

$$f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T) \quad \forall v \text{ und } \forall S \subseteq T$$

Und wozu ist das nützlich?

Wenn f zusätzlich monoton, d.h. $f(S \cup \{v\}) \geq f(S)$

dann Approximation per Hill-Climbing möglich

$\Rightarrow$ **Wenn σ submodular, dann approximierbar!**

2.1. Approximierbarkeit

Sie f eine nichtnegative,monotone submodulare, Funktion
Beginne mit einer leeren Menge S, füge jeweils nacheinander ein Element hinzu, das den größten Gewinn bringt bis k Elemente hinzugefügt.

Sei S* die k-elementige Menge, die f maximiert.
Dann f(S)>=(1-1/e)*f(S*)
(nach Nemhauser, Wolsey und Fischer)

2.2. Erweiterung

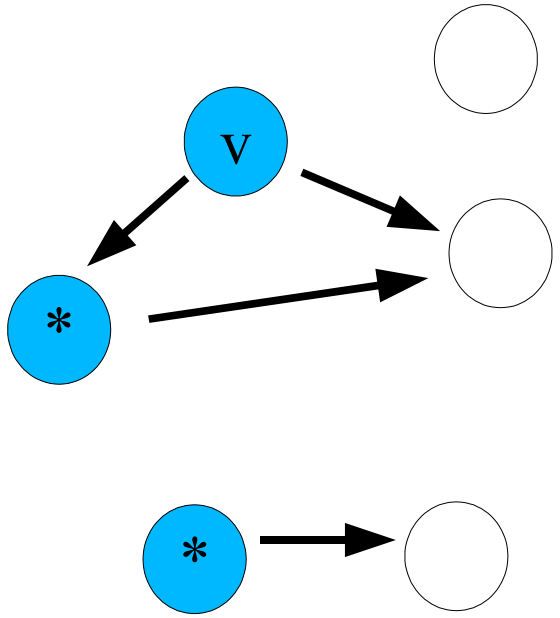
Knoten v haben Eigenschaft w_v, die angibt, wie wichtig v für Endresultat ist

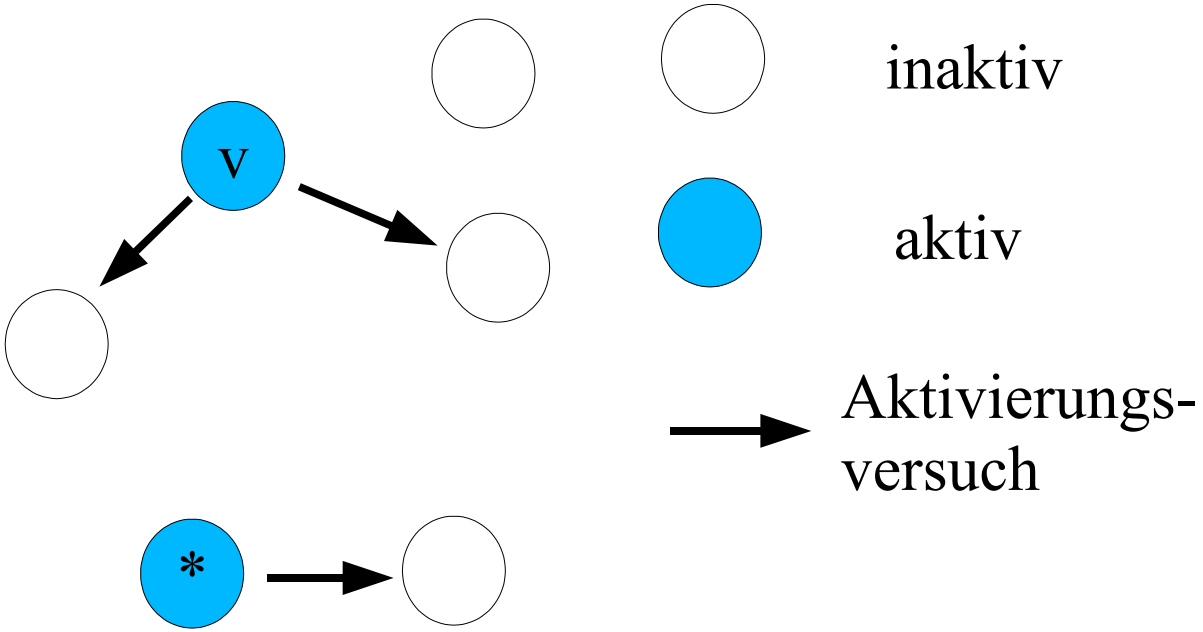
$$\sigma_w(A) = \sum_{v \in B} w_v; \quad B := \text{Am Ende aktive Knoten}$$

$$w_v = 1 \; \forall \, v \Rightarrow \text{Basisfall}$$

$\sigma_w()$ dann submodular, wenn auch $\sigma()$ submodular!

Einschub: Submodularität im Sozialen Netzwerk





*:Knoten aus T

* Knoten aus S

3.1. Allgemeines Treshold-Modell

f_v bildet Teilmenge der Nachbarn von v auf $[0,1]$ ab

$f_v(\emptyset)=0$

jeder Knoten wählt am Anfang zufällig Θ_v aus $[0,1]$

v aktiv in Schritt t, wenn $f_v(S)\geq\Theta_v$,

$S:=$ Menge der in $t-1$ aktive Nachbarn

Linear Treshold: $f_v=\sum_{u\in S} b_{v,u}$ mit $\sum_{u \text{ Nachbar von } v} b_{v,u}\leq 1$

3.2. Allgemeines Cascade-Modell

Definieren:
$p_v(u, S) \in [0,1]$ $S, \{u\}$ disjunkte Teilmengen der Nachbarn

S Menge der Nachbarn, mit gescheiterten Versuch

Independent Cascade: $p_v(u, S)$ eine von S unabhängige Konstante

3.3. Bemerkungen

- Beide Modelle sind äquivalent.
- Das Problem der Approximation ist im allgemeinen Fall NP-schwer

Wie stark kann man die Anfangsmodelle verallgemeinern, dass sie noch approximierbar bleiben?

3.4. Triggering Model

- Jeder Knoten wählt zufällig und unabhängig einen **Triggering Set** aus seinen Nachbarn aus.
- Ein Knoten wird in Schritt t aktiv, wenn ein Knoten aus seinem Triggering Set in Schritt t-1 aktiv ist

4. Nichtprogressive Prozesse

- Knoten sind abzuschaltbar
- Eingriff in laufenden Prozess möglich**!**
- Linear Treshold: Ein Knoten wählt in jedem Schritt seinen Grenzwert Θ per Zufall neu.
- Knoten v aktiv in Schritt t, wenn $f_v(S) \geq \Theta_v$
- Interventionen: Man wählt zu einer bestimmten Zeit t einen Knoten v aus um ihn zu aktivieren.

4. Nichtprogressive Prozesse

- **Influence Maximization Problem:** Finde k-Interventionen die den Einfluss maximieren!

- Über progressives Modell mit einem τ-schichtigen Graphen G^{τ} darstellbar. Wobei τ die Anzahl der Zeitschritte. =>Approximierbarkeit

5. Allgemeine Marketing Strategien.

- Knoten werden nicht direkt aktiviert
- Anzahl m an **Marketinghandlungen** Mi
- Jedes M_i kann die Wk aktiviert zu werden von einem oder mehreren Knoten erhöhen
- **Investments** x_i in Marketing Aktionen M_i

- **Markting Strategie**:= m-dimensionaler Vektor x von Investionen

5. Zu maximierende Funktion

$$g(x) = \sum_{A \subseteq V} \sigma(A) \cdot \prod_{u \in A} h_u(x) \cdot \prod_{u \notin A} (1 - h_v(x))$$

- Mit Hill-Climbing approximierbar
- $h_v(x)$ Wk das Knoten v aktiviert

Quellen:

David Kempe, Jon Kleinberg, Éva Tardos:

„Maximizing the Spread of Influence trough a sozial network“