Politechnika Wrocławska

Miary wykorzystywane w wyznaczaniu centralności sieciach społecznych

Autor: Patrycja Szwed

Zakład Systemów Informacyjnych, Politechnika Wrocławska

Wrocław, 22 listopad 2007

Agenda

- Krótkie wprowadzenie
- Przykład wprowadzający – wybór leadera
- Minimalne wymagania dla ustalenia centralności
- Miary
  1. bazujące na odległości i sąsiedztwie
  2. bazujące na najkrótszej ścieżce
  3. ogólne metody centralności krawędzi wywodzące się z definicji centralności wierzchołka
  4. ogólny zasięg aktywności miar
  5. analogia pomiędzy przepływem informacji a trendami (nurtami)
  6. bazujące na losowych procesach
  7. oceniające ważność wierzchołka poprzez ocenę centralności wierzchołków go otaczających

Miary centralności służą do określenia intuicyjnego odczucia że w większości sieci niektóre wierzchołki lub krawędzie są bardziej centralne niż inne.

Gęstość centralności (closeness centrality) – Shimbel zdefiniował *rozproszenie* jako sumę wszystkich odległości w grafie.

Przykłady wprowadzające

- Wybór leadera z grupy studentów

Dystans i sąsiedzctwo (Distances and neighborhoods)

Wyróżnia się następujące miary:

- Stopień centralności (degree) $c_D(v)$ wierzchołka –jako stopień $d(v)$ z v jeśli graf jest nieskierowany.

  W grafach skierowanych:

  - czy wewnętrzna centralność $c_{iD}(v) = d-(v)$
  - zewnętrzna centralność $c_{iD}(v) = d+(v)$

- Problem umiejscowienia lokalizacji

  1. eccentricity – miara Hage'a & Harary'ego
  2. clseness – miara

     dla grafów połączonych nieskierowanych
  3. centroid values

- Właściwości struktury

  1. centrum grafu
  2. mediana grafu
  3. centroid grafu

$$c_E(u) = \left(\frac{1}{e(u)}\right) = \frac{1}{\max\{d(u,v) : v \in V\}}$$

$$c_C(u) = \frac{1}{\sum_{v \in V} d(u,v)}$$

$$c_R(u) = \frac{\sum_{v \in V}(\Delta_G + 1 - d(u,v))}{n-1}$$

$$c_F(u) = \min\{f(u,v) : v \in V - u\}$$

Problem najkrótszej ścieżki 1/2

- Stress Centrality – mierzy liczbę połączeń które mijają element w scenariuszu wszystkie-do-wszystkich

$$C_S(v) = \sum_{s \neq v \in V} \sum_{t \neq v \in V} \sigma_{st}(v) \qquad C_S(e) = \sum_{s \in V} \sum_{t \in V} \sigma_{st}(e)$$

$$C_S(v) = \frac{1}{2} \sum_{e \in \Gamma(v)} C_S(e) - \sum_{v \neq s \in V} \sigma_{sv} - \sum_{v \neq t \in V} \sigma_{vt}$$

dla wszystkich $v \in V$

- Shortest-Path Betweenness Centrality mierzy kontrolę nad komunikacją pomiędzy innymi wierzchołkami

$$C_B(v) = \sum_{s \neq v \in V} \sum_{t \neq v \in V} \delta_{st}(v) \qquad C_B(e) = \sum_{s \in V} \sum_{t \in V} \sigma_{st}(e)$$

Problem najkrótszej ścieżki 2/2

zależność $c_B(v) = \sum_{e \in \Gamma^+(v)} c_B(e) - (n-1) = \sum_{e \in \Gamma^-(v)} c_B(e) - (n-1)$

- Traversal Sets –

$$T_e = \{(a,b) \in V \times V \exists p, p\_is\_a\_shortest\_path\_from\_a\_to\_b\_and\_contains\_e\}$$

miara dla ważności krawędzi

Miara użyta do określenia grafu z uwzględnieniem hierarchicznej organizacji.

$$c_{ts}(e) = \min\{|H| \mid H\_is\_a\_vector\_cover\_for\_T_e\}$$

Metody centralności krawędzi wywodzące się z definicji centralności wierzchołka

- Miary krawędzi pochodzące z miar wierzchołka

  - miary skierowane

  - miary nieskierowane

  - zasięg grafu

Vitality 1/2

- Vitality Index $x : V(G, x) = f(G) - f(G \setminus \{x\})$ determinują ważność wierzchołków lub krawędzi w grafie; stratę jakości grafu po usunięciu wierzchołka bądź krawędzi
- Flow Betweenness Vitality

$$c_{mf}(u) = \sum_{\substack{s,t \in V \\ u \neq s, u \neq t \\ f_{st} > 0}} \frac{f_{st}(u)}{f_{st}}$$

Vitality 2/2

- Closeness Vitality - $C_{CV}(x) = I_W(G) - I_W(G \setminus \{x\})$ mierzy ile koszty transportu w wszystkie-do-wszystkich będą wzrastać jesli korespondujący element zostanie usuniety z grafu

  średnia odległość między dwoma wierzchołkami $$d_{\varnothing}(G) = \frac{I_W(G)}{n(n-1)}$$
- Vitality-Like Index:
  - Shortcut Values
  - Stress Centrality

Current Flow

- Sieci elektryczne
- Current-Flow Betweenness Centrality

$$c_{CB}(v) = \frac{1}{(n-1)(n-2)} \sum_{s,t \in V} \tau_{st}(v)$$

Miara aktualnego przepływu części przez wierzchołek biorąc pod uwagę wszystkie możliwe pary

- Current-Flow Closeness Centrality – miara odległości pomiędzy dwoma wierzchołkami jako różnicę potencjałów $c_{CC}(v) = \frac{n-1}{\sum_{t \neq v} p_{vt}(v) - p_{vt}(t)}$

Random Processes

- Random Walks Betweenness Centrality – jest ekwiwalentna do aktualnego przepływu między wierzchołkami $c_{RWB}: V \to \Re$

$$c_{RWB}(v) = c_{CB}(v)$$

- Random-Walk Closeness Centrality

Centralność Markowa

$$c_M(v) = \frac{n}{\sum_{s \in V} m_{sv}}$$

Feedback

Centralności, gdzie węzeł jest bardziej centralny, im bardziej centralni są jego “sąsiedzi”

Grupa należąca do klasy metod grafów webowych

- Indeks Katz'a

$$c_K = \sum_{k=1}^{\infty} \alpha^k (A^T)^k 1_n$$

- Ogólne miary
  - Bonacich‘s Eigenvector Centrality
  - Hubbell Index
  - Bonacich’s Bargaining Centrality

Feedback

- Web Centralities
  - The Model of a Random-Surfer
  - PageRank
  - Algortym Hubs&Authorities
  - SALSA

Źródło:

„Centrality Indices” Dirk Koschützki, Katharina Lehmann, Leon Peeters, Stefan Richter, Dagmar Tenfelde-Podehl and Oliver Zlotowski

Dziękuję za uwagę !