Indukowana, wewnętrzna i zewnętrzna miara centralności

Barbara Rosiak

Uniwersytet Warszawski

Plan

1. Indukowana miara centralności

2. Wewnętrzna, zewnętrzna i całkowita miara centralności

3. Podsumowanie

Indukowana miara centralności

Czym jest indukowana miara centralności?

- idea konstrukcji
  1. wybrać cechę (niezmiennik) grafu
  2. usuwać wierzchołki i mierzyć zmiany w wartości niezmiennika grafu
- \+ możliwość dostosowania miary cetralności do konkretnych potrzeb
- łatwo rozszerzenie do przypadku centralności krawędziowych poprzez mierzenie wpływu usunięcia krawędzi
- łatwo rozszerzenie do przypadku centralności grupowych poprzez usuwanie zbiorów wierzchołków lub krawędzi [2] [1]

Czym jest niezmiennik grafu?

- cecha grafu
- zależy tylko od struktury grafu
- nie zależy od reprezentacji ani etykietowania grafu
- graf może być: (nie)skierowany, z wagami lub bez
- przykłady
  1. posiadanie cyklu długości k
  2. gęstość grafu $\frac{|E|}{\text{max \# możliwych krawędzi}}$
  3. przechodniość grafu $\frac{3 \cdot \text{\# przechodnich trójek}}{\text{max \# połączonych trójek wierzchołków}}$
- zazwyczaj jest liczbą, ale może być rozszerzona do zbiorów lub wektorów
  - przykład: ciąg stopni wierzchołków w grafie

Indukowana miara centralności - definicja

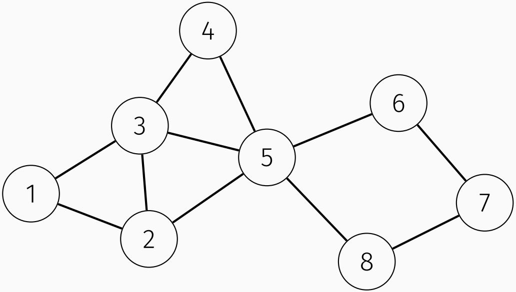
- graf G, niezmiennik grafu f, krawędź x
- indukowana miara centralności wierzchołka/krawędzi x to

$$C_f(x) = f(G) - f(G - \{x\})$$

- możemy tworzyć indukowaną miarę centalności na podstawie dowolnego niezmiennika, który jest zdefiniowany dla każdego grafu

Przykłady znanych miar centralności jako indukowane miary centralności

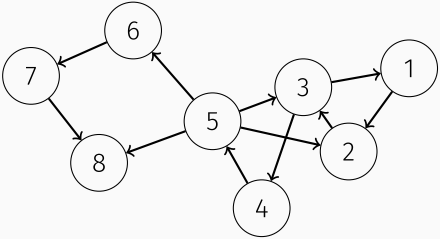
- niezmiennik: liczba krawędzi w grafie
- indukowana miara centralności: degree



Label	Degree
1	2
2	3
3	4
4	2
5	5
6	2
7	2
8	2

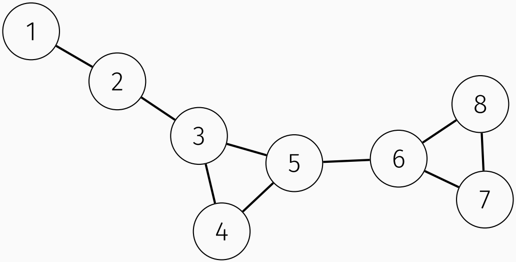
Przykłady znanych miar centralności jako indukowane miary centralności

- niezmiennik: total max flow w grafie
- indukowana miara centralności: max-flow betweenness [4] [5]



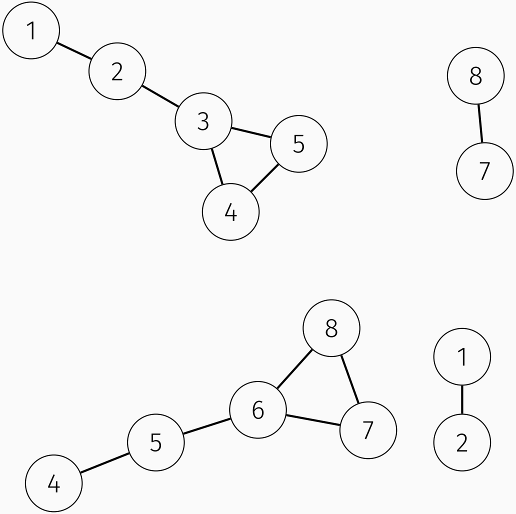
Label	Flow Betweenness
1	2
2	7
3	15
4	13
5	16
6	6
7	2
8	0

Nie wszystko można wyindukować



Label	Degree	Closeness	Betweenness
1	1	0.04348	0
2	2	0.05882	6
3	3	**0.07692**	10
4	2	0.06667	0
5	3	0.08333	12
6	3	**0.07143**	10
7	2	0.05263	0
8	2	0.05263	0

Nie wszystko można wyindukować



Pożądane cechy niezmiennika grafu

1. zdefiniowany tak, że po usunięciu wierzchołka/krawędzi niezmiennik nadal będzie dobrze zdefiniowany

   zły przykład:

   - średnica grafu i rozspójnienie grafu

2. nie powinien być normalizowany względem liczby wierzchołków/krawędzi w grafie, bo może ona usunąć wkład włożony przez usunięty element

   ekstremalny przykład:

   - gęstość grafu i pełna sieć
   - przed usunięciem gęstość grafu jest równa 1
   - po usunięciu wierzchołka gęstość grafu jest równa 1
   - podobny przykład:
     - średnia długość ścieżki w grafie
     - fragmentacja - proporcja par wierzchołków, które są dla siebie nieosiągalne do wszystkich par wierzchołków w grafie [1]

Pożądane cechy niezmiennika grafu

3. wystarczająca wrażliwość na usuwanie wierzchołków, znacząca zmiana wartości po usunięciu wierzchołka/krawędzi

   zły przykład:

   - rozmiar największej kliki w grafie
     1. mamy 2 rozłączne największe kliki, usuwamy dowolny wierzchołek
     2. mamy 1 największą klikę, usuwamy wierzchołek spoza kliki
        wierzchołki w klice mają centalność równą 1, te poza kliką 0.
   - w obu przypadkach po usunięciu wierzchołka wartość się nie zmieni

Pożądane cechy niezmiennika grafu

4. zmiana wartości niezmiennika po usunięciu wierzchołka powinna zależeć od struktury nowopowstałego grafu

   przykłady:

   1. liczba krawędzi w grafie
      - usunięcie wierzchołka powoduje usunięcie krawędzi idących do niego
      - różnica jest równa liczbie krawędzi incydentnych z usuniętym wierzchołkiem
   2. liczba przechodnich trójek w grafie
      - różnica jest równa liczbie trójek zawierający usunięty wierzchołek

Wewnętrzna, zewnętrzna i całkowita miara centralności

Wewnętrzna, zewnętrzna i całkowita miara centralności

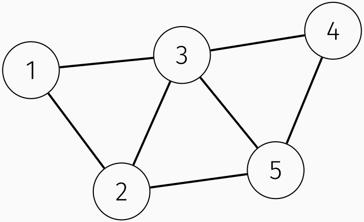
- jako niezmiennik możemy przyjąć sumę miary centralności dla każdego wierzchołka w sieci
- usunięcie wierzchołka x i obliczenie wartości na tej podstawie daje nam *całkowitą miarę centralności*

$$C_T(x) = \sum_{j \in V(G)} C(j) \quad - \sum_{j \in V(G - \{x\})} C'(j)$$

- całkowita miara centralności dla wierzchołka odzwierciedla
  - wartość centralności danego wierzchołka - wewnętrzna miara centralności
  - wpływ danego wierzchołka na wartości centralności innych wierzchołków - zewnętrzna miara centralności
- całkowita centralność = wew. centralność + zew. centralność

Przykład: Degree w grafie nieskierowanym

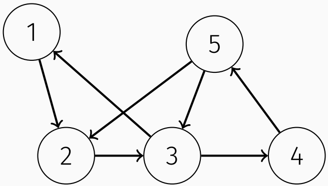
1. obliczamy stopień każdego wierzchołka - wew. centralność
2. suma stopni w całym grafie - niezmiennik
3. gdy usuniemy wierzchołek, wartość niezmiennika spadnie o dwukrotność stopnia tego wierzchołka
4. całkowita centralność odpowiada stopiowi usuniętego wierzchołka plus jego wpływie na wartość stopni jego sąsiadów



Label	Degree	Degree ($G - \{5\}$)
1	2	2
2	3	2
3	4	3
4	2	1
5	3	
sum	14	8

Przykład: In-degree w grafie skierowanym

- niezmiennik: suma in-degree dla każdego wierzchołka w grafie
- centralność całkowita: suma in-degree i out-degree danego wierzchołka
- centralność wewnętrzna: in-degree danego wierzchołka
- centralność zewnętrzna: out-degree danego wierzchołka



Przykład: Betweenness

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

gdzie σ_{st} to liczba najkrótszych ścieżek z s do t

- doodanie lub usunięcie krawędzi nieincydentnych do danego wierzchołka może zwiększyć lub zmniejszyć centralność tego wierzchołka
- tabela przedstawiająca całkowitą, wewnętrzną i zewnętrzną centralność, biorąc Betweenness jako miarę centralności wewnętrznej[3], obliczoną na podstawie sieci ślubów rodzin florenckich z okresu renesansu[6]

Przykład: Betweenness

Induced centrality scores using betweenness for endogenous centrality.

	Total centrality	Endogenous centrality	Exogenous centrality
Medici	73	48	25
Guadagni	38	23	15
Albizzi	43	19	24
Salviati	57	13	44
Ridolfi	9	10	−1
Bischeri	11	10	1
Strozzi	14	9	5
Barbadori	14	9	5
Tornabuoni	13	8	5
Castellani	18	5	13
Peruzzi	24	2	22
Pazzi	35	0	35
Ginori	28	0	28
Acciaiuoli	24	0	24
Lamberteschi	29	0	29
Pucci	0	0	0

Przykład: Reverse-closeness

- n to liczba wierzchołków w grafie
- farness - suma długości najkrótszych ścieżek między węzłem a wszystkimi innymi węzłami w grafie
- closeness - odwrotność farness
- znormalizowana closeness - closeness$\cdot(n-1)$
- usunięcie wierzchołka lub krawędzi może rozłączyć graf, co prowadzi do niezdefiniowanych odległości między parami węzłów w różnych spójnych składowych
- reverse-closeness
    - odległość między dwoma nieosiągalnymi wierzchołkami to n
    - w celu normalizacji odejmujemy sumę odległości od $n(n-1)$, maksymalnej możliwej wartości (izolowany wierzchołek, ma odległość n do pozostałych $n-1$ wierzchołków)
    - $n(n1)$ - farness, gdzie niezdefiniowane odległości są równe n

Przykład: Reverse-closeness

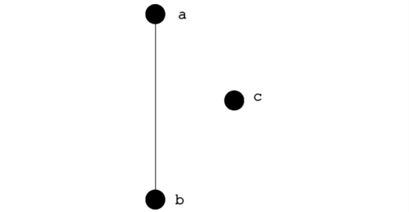


Fig. 3. The graph $P_2 \cup K_1$ which shows how an isolate can have a non-zero contribution to the non-adjusted total reverse-closeness score.

Calculating exogenous adjusted reverse-closeness centrality for node *c* in Fig. 3.

Node	Graph G		Graph G-c		
	Farness	Reverse-closeness	Farness	Simple reverse-closeness	Adjusted reverse-closeness
a	4	2	1	1	2
b	4	2	1	1	2
c	6	0	–		
Graph invariant		4	Graph invariant	2	4

- wartość niezmiennika dla początkowego grafu, $n(n-1)$−farness
- usunięcie izolowanego wierzchołka w celu policzenia całkowitej i zewnętrznej centralności
- wartość niezmiennika dla nowego grafu, $n'(n'-1)$−farness
- wpływ usuniętego wierzchołka jest równy 2

Przykład: Reverse-closeness

- intuicyjnie izolowany wierzchołek nie powinien wpływać na wartość
- maksymalna możliwa wartość farness w grafie z usuniętym wierzchołkiem jest mniejsza niż w oryginalnym
- przy obliczaniu reverse-closeness dla każdego węzła w zmniejszonym grafie powinniśmy korzystać z rozmiaru pierwotnego grafu, a nie nowego

Porównanie miar centralności

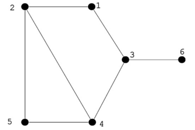


Fig. 2. Graph with a cutpoint used to demonstrate reverse-closeness calculation.

Vertex	Total centrality	Endogenous reverse-closeness	Exogenous	Betweenness
1	44	22	22	1
2	46	22	24	1.5
3	74	23	51	4.5
4	50	23	27	3
5	42	21	21	0
6	38	19	19	0

Closeness centrality scores for the Padgett Marriage Data.

	Total centrality	Endogenous centrality	Exogenous centrality	Freeman betweenness
Medici	1248	199	1049	47.5
Guadagni	702	194	508	23.2
Albizzi	692	195	497	19.3
Salviati	664	188	476	13.0
Ridolfi	402	196	206	10.3
Bischeri	398	189	209	9.5
Strozzi	392	192	200	9.3
Barbadori	396	192	204	8.5
Tornabuoni	394	195	199	8.3
Castellani	384	188	196	5.0
Peruzzi	372	186	186	2.0
Pazzi	350	175	175	0.0
Ginori	364	182	182	0.0
Acciaiuoli	372	186	186	0.0
Lamberteschi	362	181	181	0.0
Pucci	0	0	0	0.0

Twierdzenie

Twierdzenie Wierzchołek x z betweenness równym zero ma wewnętrzną miarę centralności reverse-closeness równą zewnętrznej mierze centralności reverse-closeness.

Dowód

- skierowany graf G, n-wierzchołkowy
- wierzchołek x ma betweenness równą 0
- $f(x)$ to wartość farness dla x
- T_f to suma farness w grafie
- T_f^x to suma farness w grafie po usunięciu x
- chcemy pokazać, że zewnętrzna c. == wewnętrzna c.

Twierdzenie

Dowód

- wartość całkowitej (poprawionej) reverse-closeness to $C_T(x) = 2n(n-1) + T_f^x - T_f$
- skoro wierzchołek x ma betweenness równą 0, to jego usunięcie zmniejszy farness o $2f(x)$ (usunięcie wartości farness dla x i zmniejszenie wartości wśród inncyh wierzchołków, w sumie $f(x)$)
- $T_f^x = T_f - 2f(x)$
- $C_T = 2n(n-1) - 2f(x)$
- wewnętrzna centralność $w(x) = n(n-1) - f(x)$ ((poprawiona) reverse-closeness)
- zewnętrzna c. + wewnętrzna c. = całkowita c.
- $C_T(x) = 2n(n-1) + T_f^x - T_f = 2n(n-1) + T_f - 2f(x) - T_f$
- $C_T(x) = 2n(n-1) - 2f(x) = 2(n(n-1) - f(x)) = 2w(x)$

Zależności dwustronne wierzchołków

- wcześniej patrzyliśmy na całkowitą centralność jako różnicę sum centralności w grafie z i bez usuwanego wierzchołka
- możemy również patrzeć na całkowitą centralność jako sumę różnic centralności dla każdego wierzchołka
- to podejście pozwala zdefiniować macierz p
  p_{ij} to różnica między centralnością wierzchołka j, gdy i jest w grafie i centralnością wierzchołka j, gdy i nie ma w grafie

Betweenness dependency for Fig. 2.

	1	2	3	4	5	6	Sum	Diagonal	Off-diagonal sum
1	1.0	1.5	1.5	−1.0	0.0	0.0	3.0	1.0	2.0
2	1.0	1.5	−0.5	0.0	0.0	0.0	2.0	1.5	0.5
3	1.0	−0.5	4.5	3.0	0.0	0.0	8.0	4.5	3.5
4	−3.0	−1.5	1.5	3.0	0.0	0.0	0.0	3.0	−3.0
5	0.0	1.0	1.0	2.0	0.0	0.0	4.0	0.0	4.0
6	0.5	0.0	4.0	1.5	0.0	0.0	6.0	0.0	6.0
Sum	0.5	2.0	12.0	8.5	0.0	0.0	23.0	10.0	13.0

Zależności dwustronne

- centralność wierzchołka j w grafie G ozn. $C_G(j)$, $C_{G-\{i\}}(i) = 0$

$$p_{ij} = C_G(j) - C_{G-\{i\}}(j)$$

- p_{ij} wskazuje wkład wierzchołka i w centralność wierzchołka j
- suma wiersza i daje nam wartość całkowitej centralności wierzchołka i
- p_{ii} daje nam wartość wewnętrznej centralności wierzchołka i
- suma wiersza bez $_{ii}$ daje nam wartość zewnętrznej centralności wierzchołka i
- macierz tę można interpretować jako macierz zależności wierzchołków - p_{ij} to stopień, w jakim wierzchołek j zależy na wierzchołku i dla centralności

Podsumowanie

Podsumowanie

1. ogólna metoda konstruowania miar centralności wykorzystująca niemal dowolny niezmiennik grafu
   - oblicz wartość niezmiennika
   - usuń wierzchołek
   - oblicz nową wartość niezmiennika
   - oblicz różnicę tych wartości
2. interpretacja indukowanej miary centralności jako wkład danego węzła w daną cechę grafu
3. naturalna generalizacja centralności węzła do centralności krawędzi lub grupowej
4. całkowita centralność, można skonstruować miarę centralności indukowanej praktycznie z dowolnej innej miary centralności
   - konstruowanie miary centralności indukowanej praktycznie z dowolnej innej miary centralności
   - suma wszystkich wyników centralności jako niezmiennik dla centralności indukowanej
   - mierzy wkład węzła do sumy centralności w sieci
   - centralność wewnętrzna i zewnętrzna

Dziękuję za uwagę!

References i

S. Borgatti.
Identifying sets of key players in a social network.
Computational Mathematical Organization Theory, 12:21–34, 04 2006.

M. Everett and S. Borgatti.
The centrality of groups and classes.
Journal of Mathematical Sociology, 23:181–201, 01 1999.

M. G. Everett and S. P. Borgatti.
Induced, endogenous and exogenous centrality.
Social Networks, 32(4):339–344, 2010.

References ii

L. Freeman, S. Borgatti, and D. White.
Centrality in valued graphs: A measure of betweenness based on network flow.
Social Networks, 13:141–154, 06 1991.

Koschützki, Dirk, K. Zweig, K. A., Peeters, Leon, Richter, S. Richter, Tenfelde-Podehl, Dagmar, Zlotowski, and Oliver.
***Centrality Indices*, volume 3418, pages 16–.**
01 2005.

J. Padgett and C. Ansell.
Robust action and the rise of the medici, 1400-1434.
American Journal of Sociology - AMER J SOCIOL, 98, 05 1993.