

Analiza Sieci Społecznych

Pajek

Dominik Batorski*
Instytut Socjologii UW

3 czerwca 2016

1 Wprowadzenie

Regularności we wzorach relacji często są nazywane strukturą. Analiza sieci społecznych jest zbiorem metod służących badaniu tych struktur (Knoke i Kuklinski, 1982; Wasserman i Faust, 1994; Hanneman).

Badając zjawiska społeczne można obserwować zarówno: **cechy jednostek i relacje między nimi**. O grupie należy myśleć w kategoriach zależności pomiędzy cechami jednostek i w kategoriach relacji między jednostkami (same atrybuty nie wystarczą). Badania społeczne wymagają uwzględnienia obydwu możliwych podejść - zarówno atrybutów jednostek jak i kontekstu, w którym się znajdują. Umożliwiają to właśnie analizy sieci.

Sieć to osoby lub obiekty powiązane relacjami.

2 Pajek

Pajek jest programem do analizy i graficznej prezentacji sieci. Nadaje się do analizy stosunkowo dużych sieci. Jest szczególnie przydatny do wstępnych, eksploracyjnych analiz i wizualizacji (Batagelj i Mrvar 1998). Program jest darmowy i można go ściągnąć ze strony: <http://mrvar.fdv.uni-lj.si/pajek/>

Na stronie tej można znaleźć również Manual z opisem programu, jego możliwości i sposobów wykorzystania go, a także liczne zbiory danych, które można samodzielnie analizować.

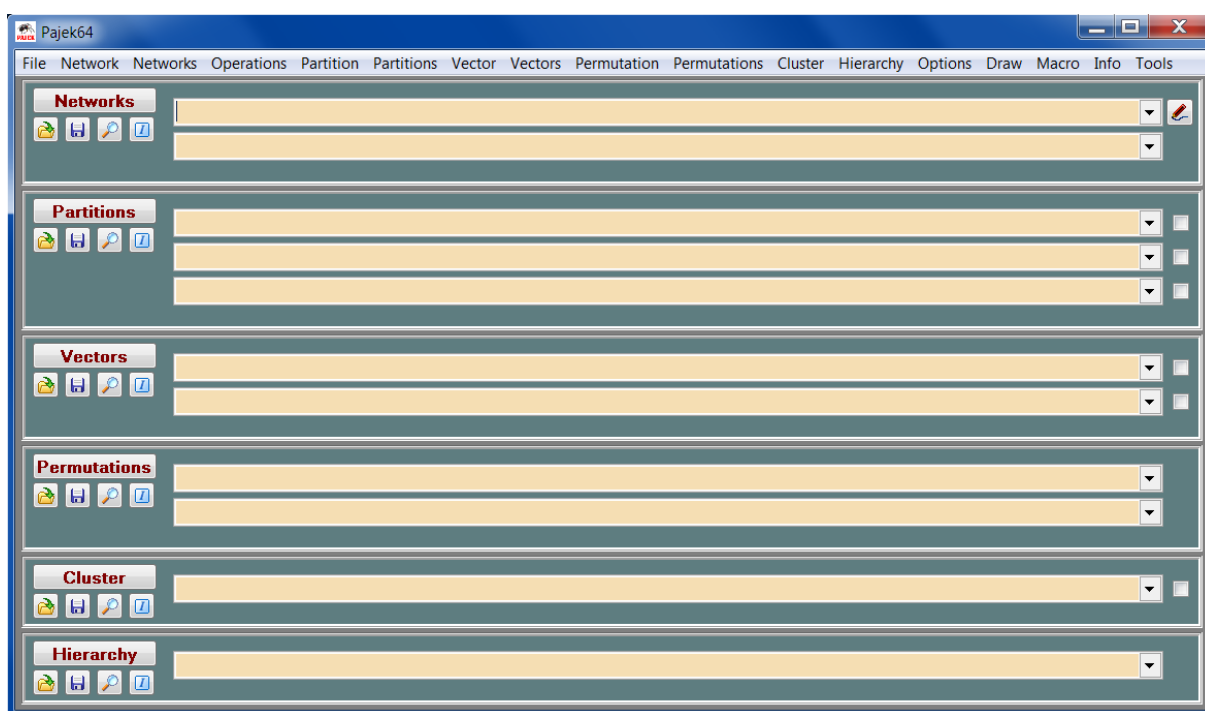
3 Dane

Główne okienko programu Pajek zawiera sześć różnych struktur danych:

*e-mail: db@uw.edu.pl WWW: <http://stat.is.uw.edu.pl/db/>

Struktura	Opis	Rozszerzenie
network	sieć – węzły i połączenia	.NET
partition	węzły i przynależność do grup	.CLU
vector	wektor wartości dla węzłów	.VEC
permutation	uporządkowanie węzłów	.PER
cluster	podzbiory węzłów	.CLS
hierarchy	hierarchicznie uporządkowane clustry i węzły	.HIE

Tabela 1: Lista rodzajów danych używanych przez program Pajek.



Rysunek 1: Główne okienko programu Pajek.

3.1 Sieć

Dane o sieci są zapisywane w pliku ASCII lub UTF8 z BOM (rozszerzenie .NET). Plik może być stworzony w zwykłym notatniku lub dowolnym edytorze tekstu.

Lista par

Format danych, przykład:

```

*Vertices n
węzeł1 "nazwa" x y z kształt [dodatkowe parametry]
...
*Arcs
węzeł1 węzeł2 siłapołączenia
...
*Edges
węzeł1 węzeł2 siłapołączenia
...

```

Lista sąsiadów

Nadaje się wyłącznie do danych, w których relacje nie mają wagi.

```

*Vertices n
węzeł1 "nazwa" x y z kształt [dodatkowe parametry]
...
*Arcslist
węzeł1 węzeł2 węzeł5
...
*Edgeslist
węzeł1 węzeł3 węzeł6
...

```

Macierz

```

*Vertices n
węzeł1 "nazwa" x y z kształt [dodatkowe parametry]
...
*Matrix
0 1 0 1 0 1
...

```

Sieci afiliacji

```

*Vertices n k
węzeł1 "nazwa" x y z kształt [dodatkowe parametry]
...
*Arcs
węzeł1 węzeł2
...
*Edges
węzeł1 węzeł2
...

```

Ważne jest by węzły jednego rodzaju były umieszczone obok siebie.

3.2 Partition - podziały sieci

Dane dotyczące sieci zawarte w pliku `.net` dostarczają informacji jedynie o węzłach i relacjach pomiędzy nimi (ewentualnie również o atrybutach tych relacji). Jednak w analizie sieci społecznych bardzo często istotne są również informacje o atrybutach węzłów (np. płci osób). W programie **Pajek** tego typu dane dotyczące podziału węzłów na podgrupy (kategorie) trzymane są w plikach o rozszerzeniu `.clu` i nazywane są *partition*. Proszę zauważyć, że podział węzłów w sieci na podgrupy może wynikać zarówno z własności ich pozycji strukturalnej, jak i z charakterystyk niezwiązanych z pozycją (np. wspomniana płeć).

```
*Vertices n
NrGrupyWęzła1
NrGrupyWęzła2
...
```

3.3 Vector - atrybuty węzłów

Rozszerzenie `.vec` *vector*. W przeciwieństwie do podziałów sieci, wartości przypisywane jednostkom w sieci mogą mieć wartości, które nie są liczbami całkowitymi.

```
*Vertices n
WartośćDlaWęzła1
WartośćDlaWęzła2
...
```

3.4 Projekt

Pajek pozwala również na zapisanie wszystkich struktur danych jako jednego projektu i jednego pliku (o rozszerzeniu `.paj`). Jest to wygodna metoda, jeśli pracujemy używając wielu różnych obiektów.

4 Operacje na danych

Operacje na danych wykonywane są poprzez opcje dostępne w menu głównego okienka programu.

- Zakładka **File** pozwala otwierać i zapisywać dane. W niektórych przypadkach możliwe jest również otwarcie trybu edycji danych. Umożliwia on dokonanie modyfikacji danych i tworzenie nowych (sieci, podziałów, itd.).
- Zakładka **Network** służy do wykonywania operacji na sieci.
- Zakładka **Networks** w odróżnieniu od **Network** umożliwia dokonywanie operacji na dwóch lub większej liczbie sieci.

- Zakładka **Operations** pozwala dokonywać operacji wymagających użycia obiektów różnego rodzaju - na przykład sieci i podziału lub wektora.
- Zakładka **Partition** służy do wykonywania operacji na podziałach sieci.
- Zakładka **Partitions** umożliwia operacje na więcej niż jednym podziale.
- Zakładka **Vector**
- Zakładka **Vectors**
- Zakładka **Permutation**
- Zakładka **Permutations**
- Zakładka **Cluster**
- Zakładka **Hierarchy**

Pozostałe zakładki z głównego menu (**Options, Draw, Macro, Info, Tools**) nie służą operacjom na danych i zostaną omówione później.

4.1 Operacje na węzłach i relacjach

Operacje na węzłach i relacjach w jednej sieci można przeprowadzać poprzez zakładkę **Network** → **Create New Network** → **Transform**.

usuwanie wybranych relacji:

Network → **Create New Network** → **Transform** → **Remove**

pozwała usunąć relacje: skierowane; nieskierowane; połączenia wielokrotne; pętle (relacje węzła z samym sobą); połączenia, których wartości spełniają zadane kryteria.

zmiana relacji nieskierowanych w skierowane:

Network → **Create New Network** → **Transform** → **Edges→Arcs**

zmiana relacji skierowanych w nieskierowane:

Network → **Create New Network** → **Transform** → **Arcs→Edges**

operacje na wartościach relacji:

Network → **Create New Network** → **Transform** → **Line values**

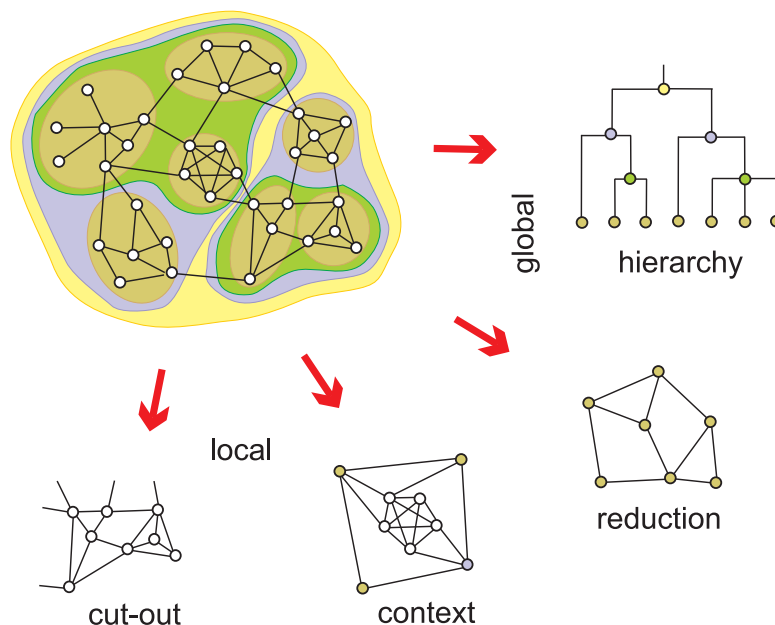
zmiana sieci dwumodalnej w jednomodalną:

Network → **2-Mode Network** → **2-Mode to 1-Mode**

Są też inne zakładki.

5 Sposoby postępowania z dużymi sieciami

Pajek jest programem, który najlepiej nadaje się do eksploracji i wizualizacji sieci. Jego zaletą jest też to, że stosunkowo dobrze radzi sobie z dużymi sieciami – mającymi kilkaset lub kilka tysięcy węzłów¹.



Rysunek 2: Sposoby postępowania z dużymi sieciami. Rysunek pochodzi z *Pajek Reference Manual. List of commands with short explanation*.

Możliwe sposoby postępowania z dużymi sieciami to:

- redukcja
- kontekst
- hierarchie
- wycinki

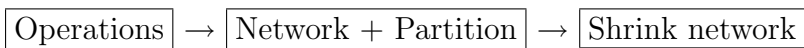
5.1 Redukcja

`Operations` → `Network + Partition` → `Shrink network`

Ważne aby podać, że wszystkie komponenty mają być zredukowane.

¹Twórcy programu sugerują, że możliwe jest analizowanie za jego pomocą sieci, które mają kilka milionów węzłów. Jednak z moich doświadczeń wynika, że działa to bardzo słabo i lepiej jest napisać własny program do analizy dużych danych.

5.2 Kontekst

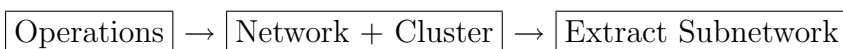
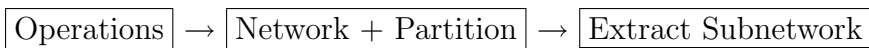


Należy zaznaczyć, który z komponentów ma nie zostać zredukowany.

5.3 Hierarchie



5.4 Wycinki



Literatura

1. Batagelj, V. i Mrvar, A. (1998). Pajek program for large network analysis. *Connections*, 21 (2), 47-57.
2. de Nooy, W., Mrvar, A., i Batagelj, V. (2011). *Exploratory Social Network Analysis with Pajek: Revised and Expanded. 2nd Edition*. New York: Cambridge University Press.
3. Hanneman, R. *Introduction to social network methods*.
4. Knoke, D. i Kuklinski, J. (1982). *Network analysis*. Sage.
5. Mrvar, A. i Batagelj, V. (2016). Analysis and visualization of large networks with program package Pajek. *Complex Adaptive Systems Modeling*, 4:6.
<http://casmodeling.springeropen.com/articles/10.1186/s40294-016-0017-8>
6. Wasserman, S. i Faust, K. (1994). *Social networks analysis: Methods and applications*. Cambridge University Press.