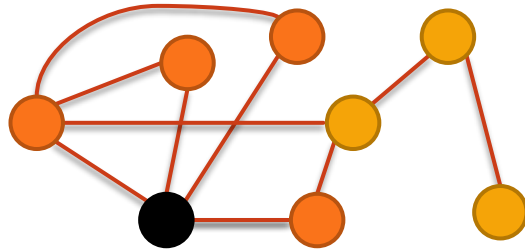


Sieci złożone

ANALIZA ZGRUPOWAŃ

Gronowanie

► Sieć ego



N – liczba sąsiadów w sieci ego

wszystkich możliwych połączeń między sąsiadami $\frac{N(N-1)}{2}$

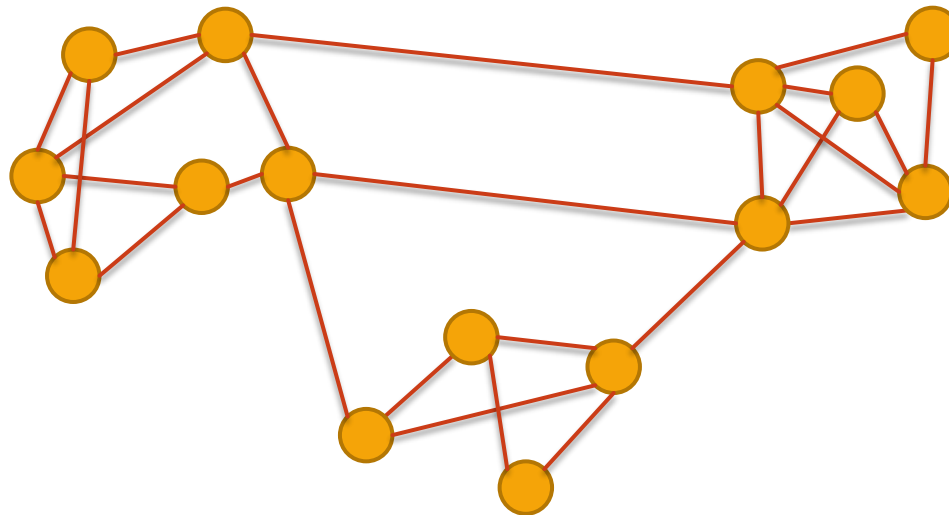
faktycznie istniejących połączeń między sąsiadami $|\mathcal{L}|$

► Współczynnik gronowania (ang. *clustering coefficient*, *clicquishness*) w grafie prostym

- gęstość sieci ego $\frac{2|\mathcal{L}|}{N(N-1)}$ sieć pusta: 0, sieć pełna: 1
- w socjologii: 3 * liczba trójkątów w sieci / liczba dróg o długości 2 w sieci
- prawdopodobieństwo, że **sąsiedzi danego wężła, są również swoimi sąsiadami**
- skłonność do utworzenia bezpośredniego połączenia przez parę węzłów mających wspólnego sąsiada

Moduły

- ▶ Kliki dobrze opisują sieci gęste, tj. $|\mathcal{L}| \sim N^2$
- ▶ W sieciach rzeczywistych $|\mathcal{L}| \sim N$
- ▶ Gęstość połączeń w obrębie modułu $\gg$ gęstość połączeń międzymodułowych



Moduły

- ▶ Problem: jak ocenić liczbę modułów obecnych w sieci?
- ▶ Analiza hierarchii skupień

- ▶ 0. Definiujemy odległość między węzłami

- ▶ $d(v_i, v_j)$

- ▶ **odległość strukturalna** (ile jest tych samych sąsiadów)

- ▶ 1. Znajdujemy najbliższą parę węzłów lub modułów

- ▶ 2. Łączymy je w jeden moduł

- ▶ 3. Przeliczamy macierz odległości D_{ij}

- ▶ Wracamy do kroku 1, dopóki wszystko stanie się jednym modułem.

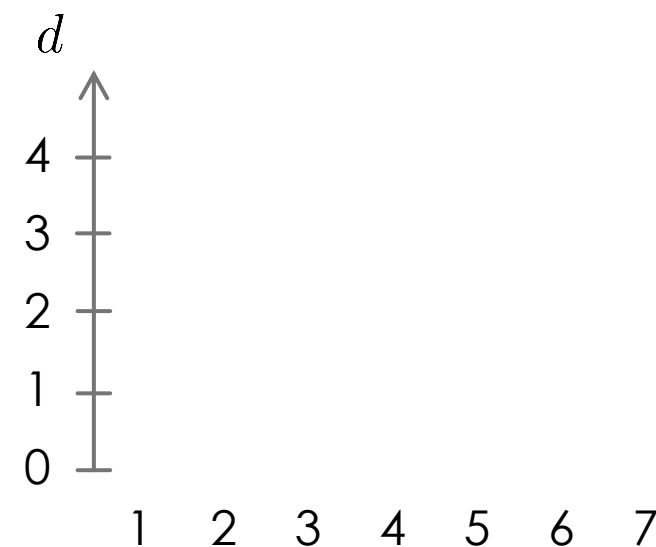
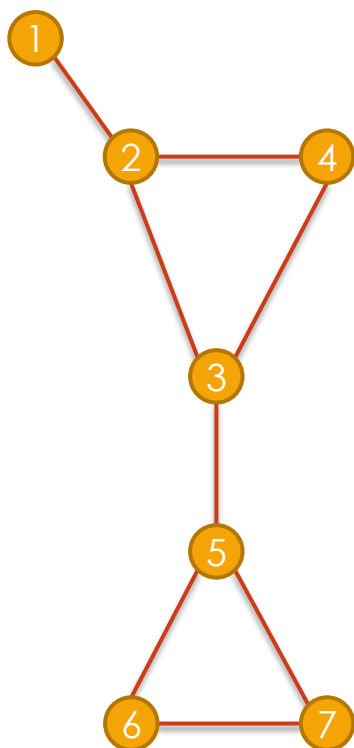
$$D_{ij} = \left[d(v_i, v_j) \right]$$

$$D_{ij} = \sqrt{\sum_{k=1}^N (a_{ik} - a_{jk})^2}, \quad k \neq i, j$$

Moduły

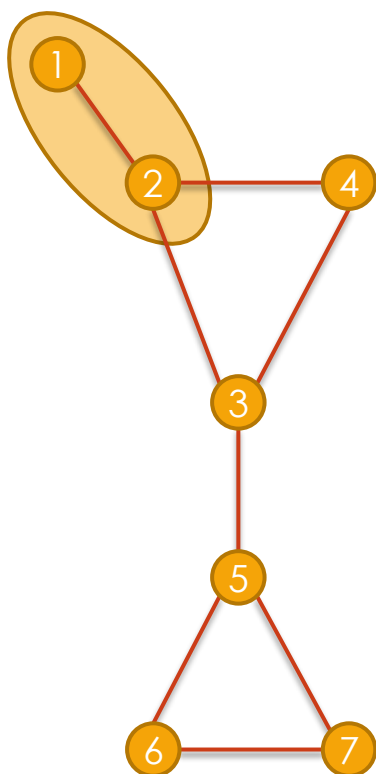
- ▶ Jak wyznaczyć **odległość między modułami**?
 - ▶ najbliższe sąsiedztwo: najkrótsza odległość między elementami obu modułów
 - ▶ najdalsze sąsiedztwo,
 - ▶ średnie sąsiedztwo.

Przykład

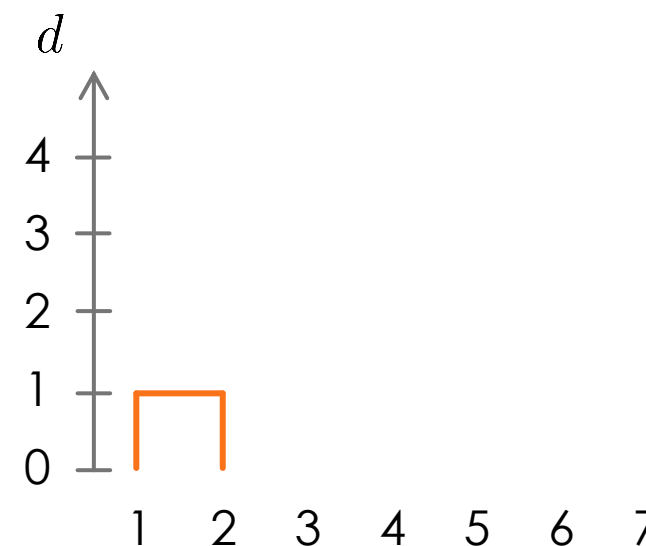


Dendrogram

Przykład

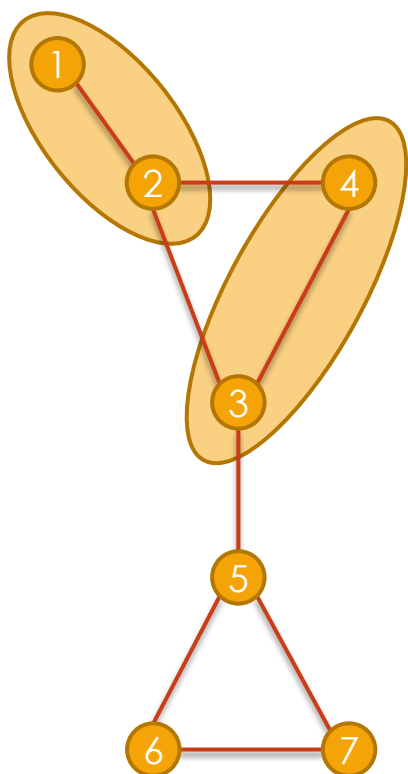


Odległość między modułami:
najdalszy sąsiad

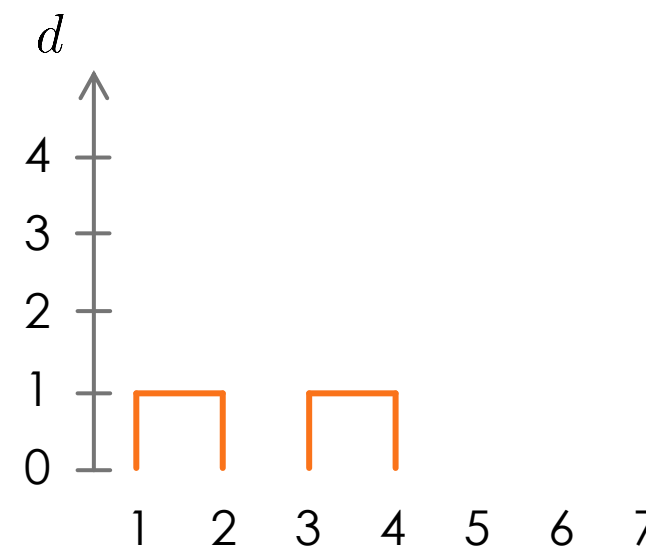


Dendrogram

Przykład



Odległość między modułami:
najdalszy sąsiad

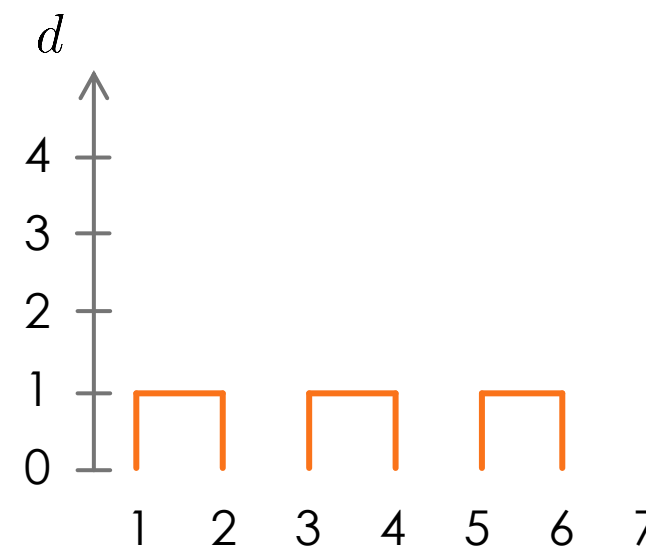
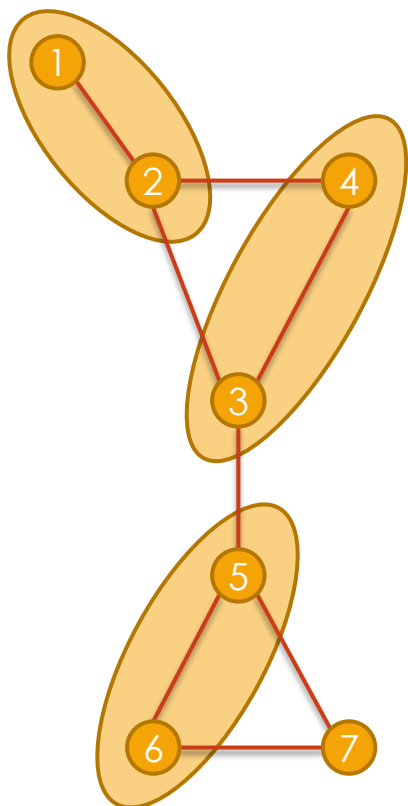


Dendrogram

Przykład



Odległość między modułami:
najdalszy sąsiad

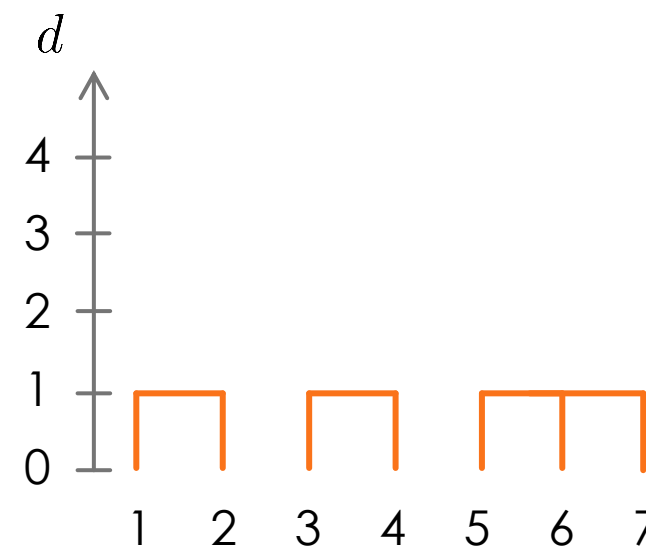
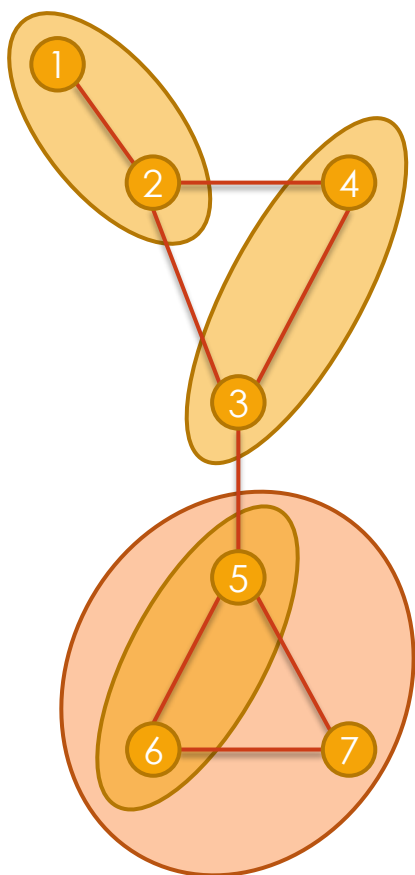


Dendrogram

Przykład



Odległość między modułami:
najdalszy sąsiad

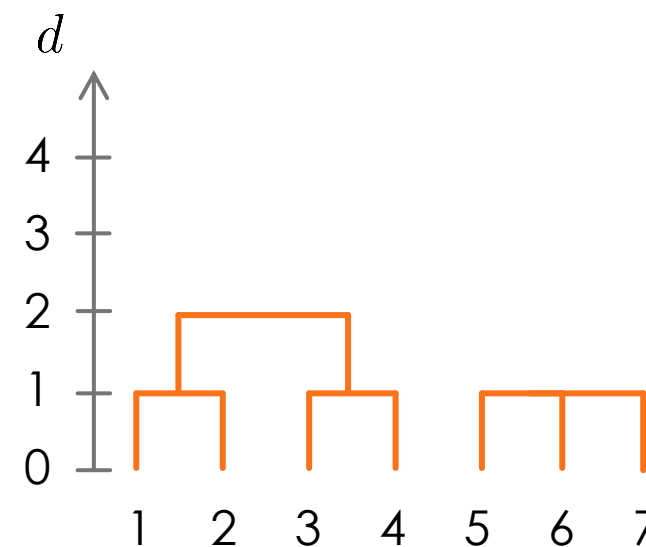
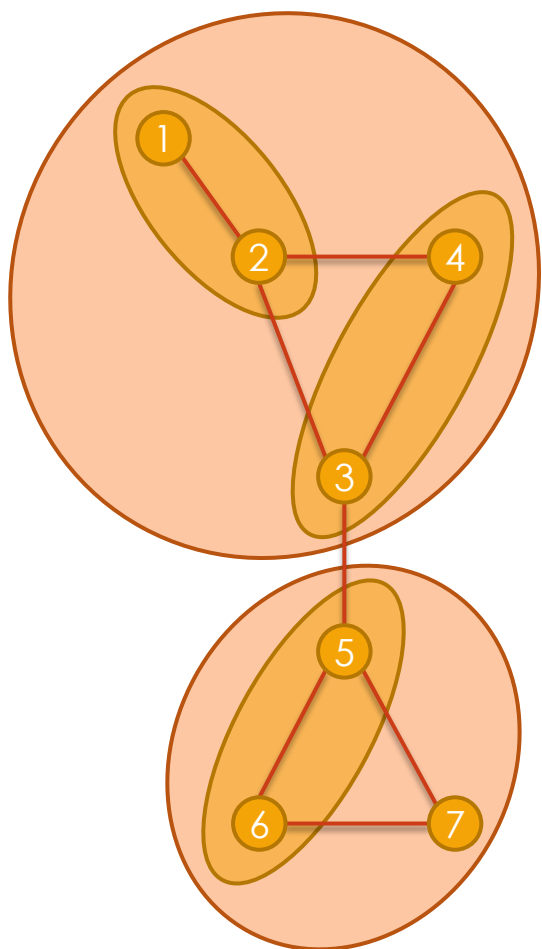


Dendrogram

Przykład



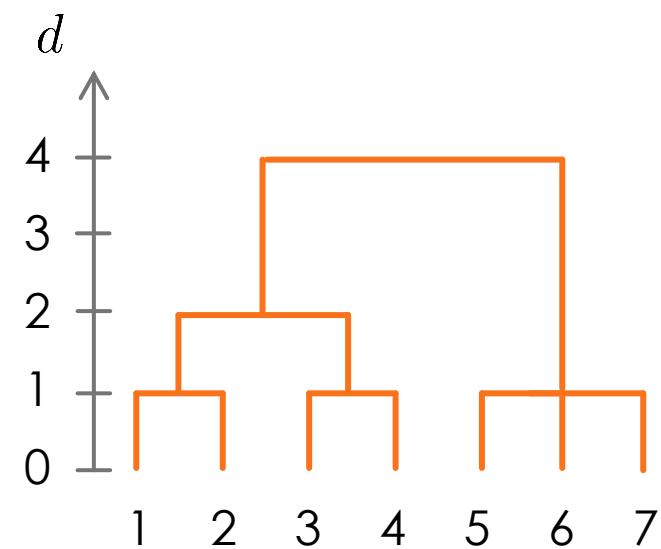
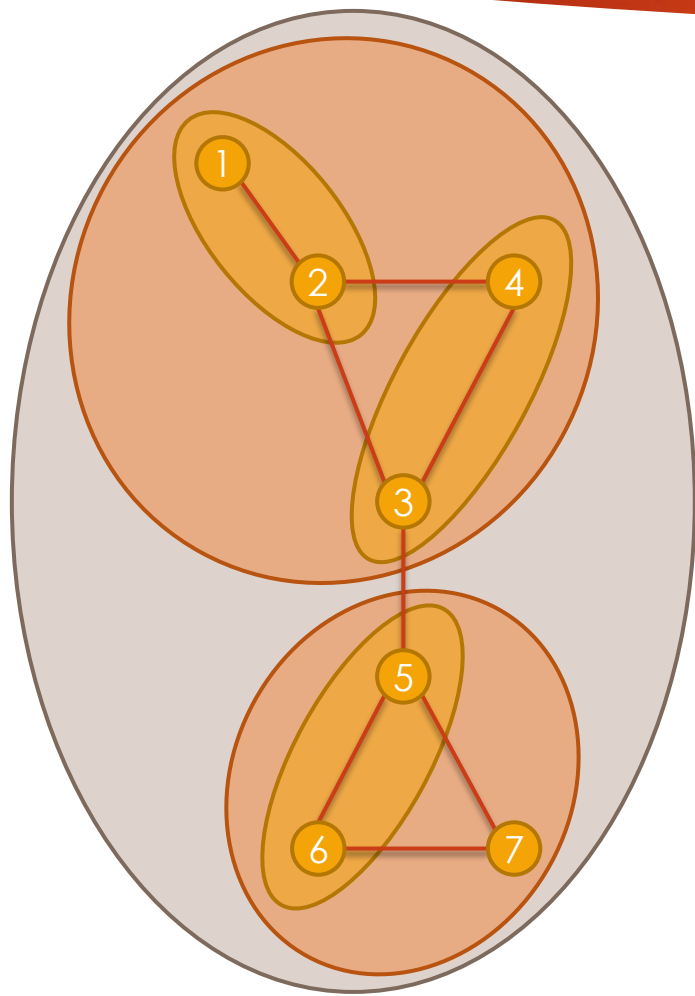
Odległość między modułami:
najdalszy sąsiad



Dendrogram

Przykład

Odległość między modułami:
najdalszy sąsiad

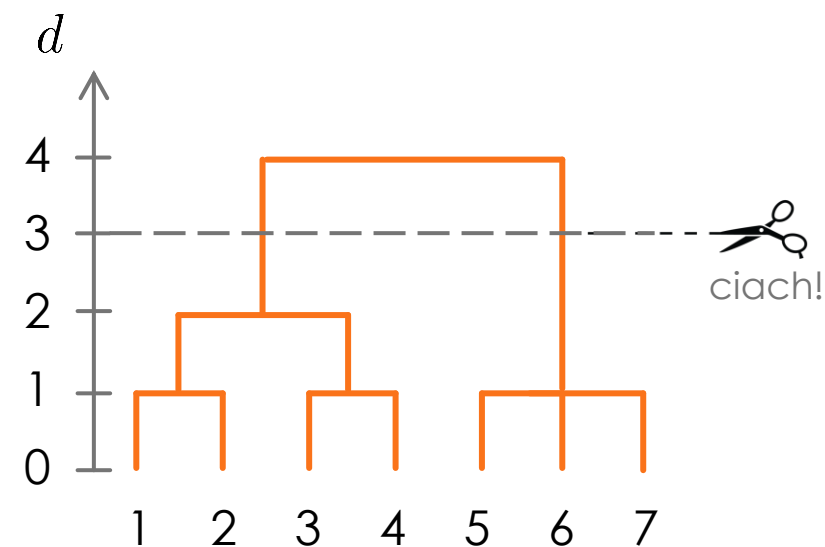
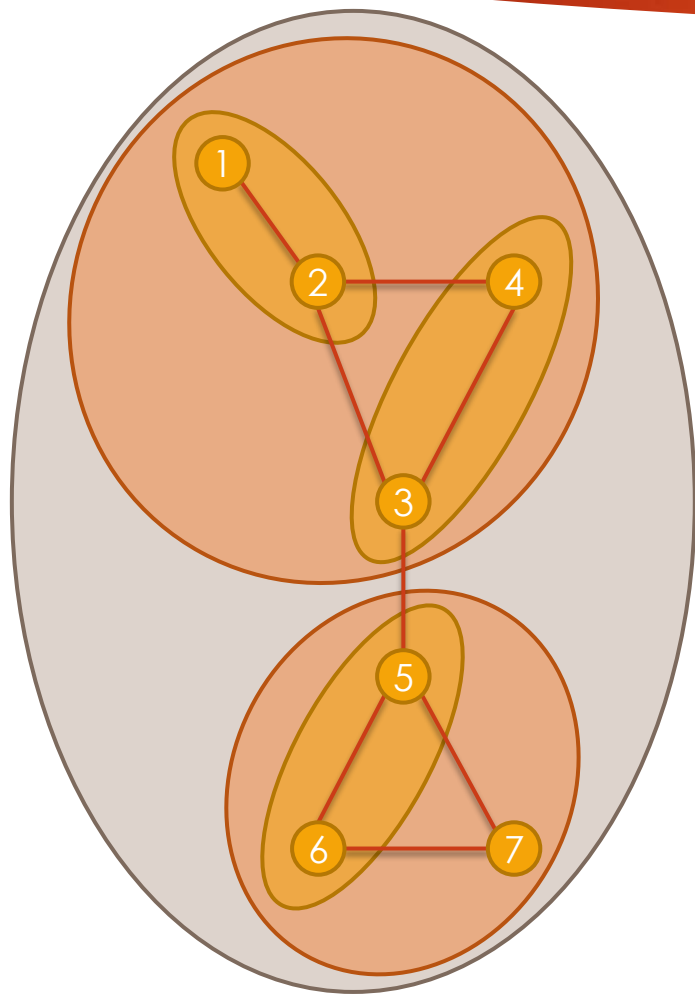


Dendrogram



Przykład

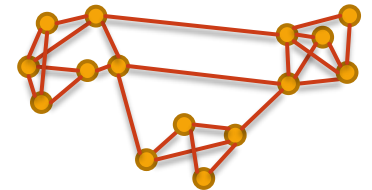
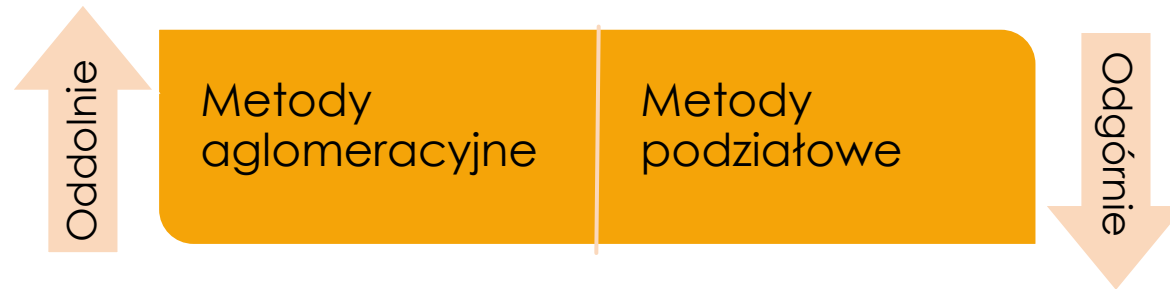
Odległość między modułami:
najdalszy sąsiad



Dendrogram



Metody aglomeracyjne i podziałowe

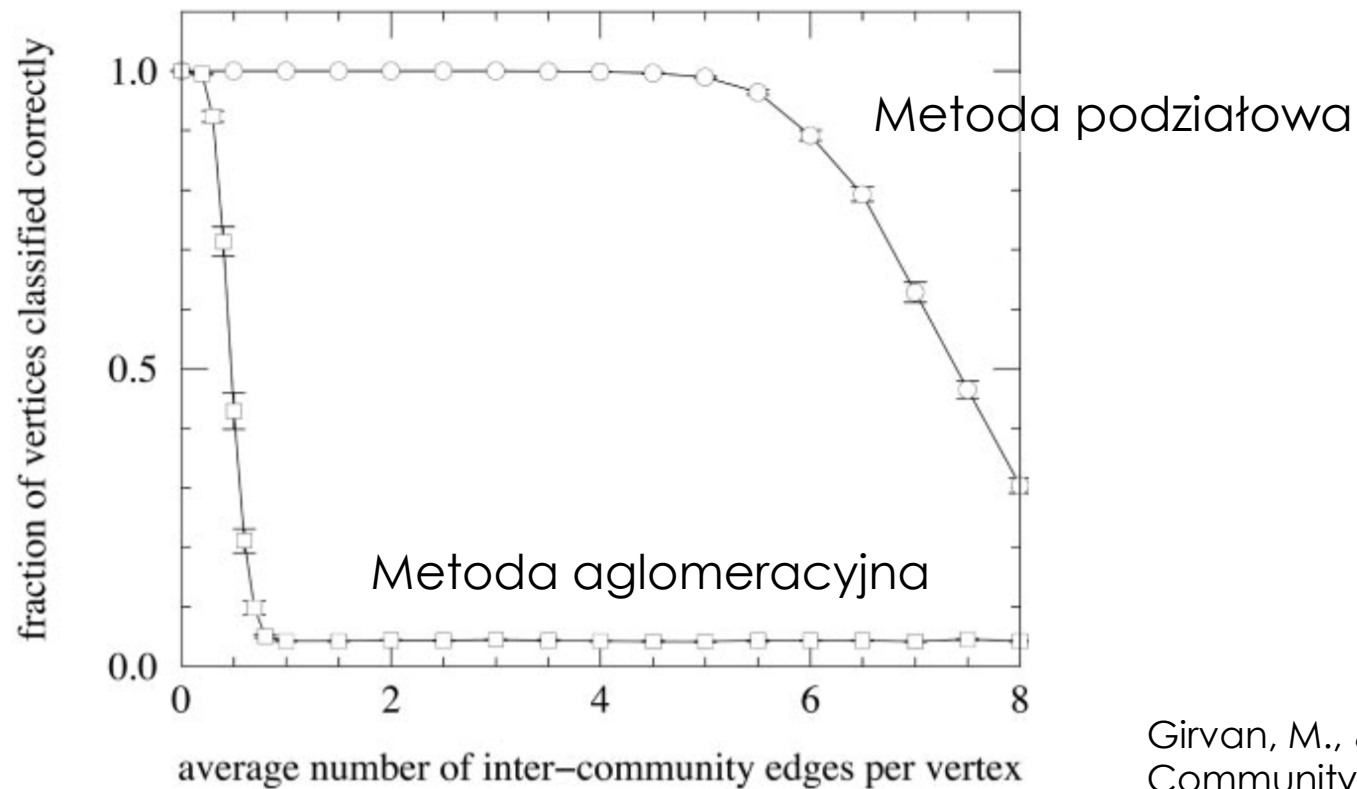


- ▶ Zamiast odległości analizuje się **pośrednictwo**
 - ▶ Łączy dwóch modułów powinny mieć dużą wartość pośrednictwa
 - ▶ 1. Znajdujemy połączenia o największym pośrednictwie i usuwamy je z sieci
 - ▶ 2. Jeżeli sieć się nie rozspójniła na oddzielne moduły, powtarzaj krok 1
 - ▶ Wracamy do kroku 1, dopóki znikną wszystkie połączenia (każdy węzeł stanie się modułem)

Porównanie metod na **sieci symulowanej**

- ▶ Wybierz liczbę węzłów
 - ▶ Wybierz liczbę modułów, które mają pojawić się w generowanej sieci
 - ▶ Przydziel węzły do modułów
 - ▶ Węzły w obrębie tego samego modułu łącz z prawdopodobieństwem p_w
 - ▶ Węzły z różnych modułów łącz z prawdopodobieństwem p_m
-
- ▶ Na sieciach o różnych konfiguracjach wartości p_w, p_m przeprowadź grupowanie metodą podziałową i aglomeracyjną
 - ▶ Porównaj wyniki i wyciągnij wnioski z różnic między metodami

Porównanie metod na sieci symulowanej



Girvan, M., & Newman, M. E. (2002).
Community structure in social and biological networks.
Proceedings of the national academy of sciences, 99(12), 7821-7826.