Detecting communities in directed acyclic networks

Suzana Antunović [1] Damir Vukičević [2]

[1]Faculty of Civil Engineering, Architecture and Geodesy

[2]Faculty of Natural Sciences

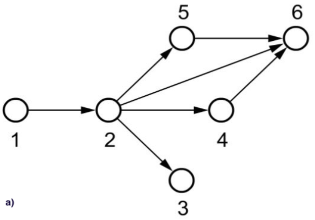
2nd Croatian Combinatorial Days

Basic idea

- If G is a network with topologically ordered vertices $x_1 \prec x_2 \prec ... \prec x_n$, we seek division of a network into communities $A_1, A_2, ..., A_k$ in such a way that:

  if $x_i \prec x_j$, $x_i \in A_p$ and $x_j \in A_q$ then $A_p \prec A_q$ or $A_p = A_q$.

Goals



a) Simple example of a network with $n = 6$ vertices i $m = 7$ directed edges. **b)** Division into 2 consecutive communities after applying the algorithm

Challenges

- formulation of the term *"community"*

Challenges

- formulation of the term *”community”*
- community detection in *<u>directed</u>* networks

Challenges

- formulation of the term *"community"*
- community detection in <u>*directed*</u> networks
- apart from the edge direction, the requirement for *topological sort* must be considered

Algorithm

Algorithm 1 Algorithm for consecutive community detection

each vertex is assigned with unique numerical label $l_i \in \{1, 2, ..., n\}$
vertices are placed in ascending order
while there are vertices which haven't been considered **do**
    for each vertex k starting with the last **do**
        calculate modularity change for each case $z_{k(k+1)...n}$, $z_{k(k+1)...(n-1)} + r_n$, $z_{k(k+1)...(n-2)} + r_{(n-1)}$, $\cdots$, $z_k + r_{(k+1)}$
        specify the optimal solution r_k
        place vertex k into the appropriate community in accordance with the solution obtained
    end for
end while

Algorithm

$\boxed{i=6}$ $\boxed{Z_6}$ $[6] \to r_6$

$\boxed{i=5}$ $\boxed{Z_{56}}$ $[5\ 6] \to r_5$

$Z_5 + r_6$ $[5]\ [6]$

$\boxed{i=4}$ $\boxed{Z_{456}}$ $[4\ 5\ 6] \to r_4$

$Z_{45} + r_6$ $[4\ 5][6]$

$Z_4 + r_5$ $[4]\ [5\ 6]$

$\boxed{i=3}$ Z_{3456} $[3\ 4\ 5\ 6]$

$Z_{345} + r_6$ $[3\ 4\ 5]\ [6]$

$Z_{34} + r_5$ $[3\ 4]\ [5\ 6]$

$\boxed{Z_3 + r_4}$ $[3]\ [4\ 5\ 6] \to r_3$

$\boxed{i=2}$ Z_{23456} $[2\ 3\ 4\ 5\ 6]$

$Z_{2345} + r_6$ $[2\ 3\ 4\ 5]\ [6]$

$Z_{234} + r_5$ $[2\ 3\ 4]\ [5\ 6]$

$\boxed{Z_{23} + r_4}$ $[2\ 3]\ [4\ 5\ 6] \to r_2$

$Z_2 + r_3$ $[2]\ [3]\ [4\ 5\ 6]$

$\boxed{i=1}$ Z_{123456} $[1\ 2\ 3\ 4\ 5\ 6]$

$Z_{12345} + r_6$ $[1\ 2\ 3\ 4\ 5]\ [6]$

$Z_{1234} + r_5$ $[1\ 2\ 3\ 4]\ [5\ 6]$

$\boxed{Z_{123} + r_4}$ $[1\ 2\ 3]\ [4\ 5\ 6] \to r_1$

$Z_{12} + r_3$ $[1\ 2]\ [3]\ [4\ 5\ 6]$

$Z_1 + r_2$ $[1]\ [2\ 3]\ [4\ 5\ 6]$

Complexity

- To put the vertex k in the appropriate community the algorithm considers $n-k+1$ cases
- For each case, the change in modularity is calculated.
- If we denote $d_k = d^{in}(k) + d^{out}(k)$, to correctly assign vertex k it takes $(n-k+1)d_k$ operations
- Sum through all the vertices gives the total complexity of the algorithm is $O(nm)$

Evaluation

- Curriculum networks

- Directed modularity

$$Q_d = \frac{1}{m} \sum_{1 \leq i,j \leq n} \left[A_{ij} - \frac{d^{in}(j) d^{out}(i)}{m} \right] \delta(l_i, l_j)$$

Evaluation

			Stručnjak		AORZ	
	n	m	Q_d	N_c	Q_d	N_c
Skup $\mathbb{Q}$	47	254	0.311	5	0.377	4
Elementarne funkcije	84	502	0.239	6	0.286	8
Integral	223	655	0.455	10	0.484	10
Obrada pod.	54	197	0.389	6	0.430	6
Model primarne proizvodnje	28	93	0.237	3	0.259	3
Fizika	31	49	0.238	6	0.375	4

The end

Thank you for your attention!