Σύνθετα Δίκτυα

com+plex: with+ -fold (having parts)

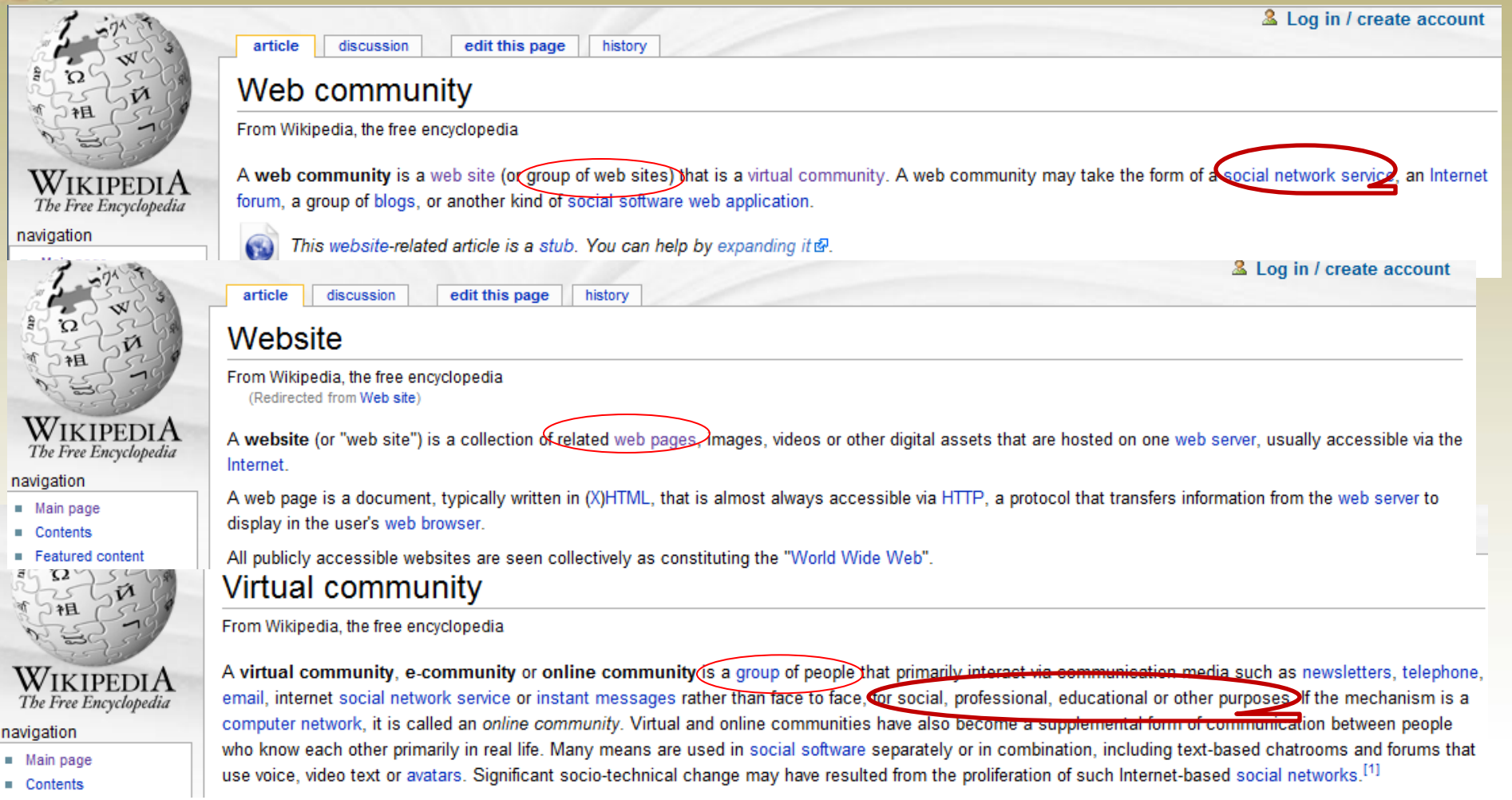
Διδάσκων –
Δημήτριος Κατσαρός

Διάλεξη 09η

Communities in Complex Networks

Κοινότητες σε Σύνθετα Δίκτυα

Τι είναι μια Web community;



Log in / create account

article | discussion | edit this page | history

Web community

From Wikipedia, the free encyclopedia

A **web community** is a web site (or group of web sites) that is a virtual community. A web community may take the form of a social network service, an Internet forum, a group of blogs, or another kind of social software web application.

This website-related article is a stub. You can help by expanding it.

navigation

Log in / create account

article | discussion | edit this page | history

Website

From Wikipedia, the free encyclopedia
(Redirected from Web site)

A **website** (or "web site") is a collection of related web pages, images, videos or other digital assets that are hosted on one web server, usually accessible via the Internet.

A web page is a document, typically written in (X)HTML, that is almost always accessible via HTTP, a protocol that transfers information from the web server to display in the user's web browser.

All publicly accessible websites are seen collectively as constituting the "World Wide Web".

navigation

- Main page
- Contents
- Featured content

Virtual community

From Wikipedia, the free encyclopedia

A **virtual community**, **e-community** or **online community** is a group of people that primarily interact via communication media such as newsletters, telephone, email, internet social network service or instant messages rather than face to face, for social, professional, educational or other purposes. If the mechanism is a computer network, it is called an *online community*. Virtual and online communities have also become a supplemental form of communication between people who know each other primarily in real life. Many means are used in social software separately or in combination, including text-based chatrooms and forums that use voice, video text or avatars. Significant socio-technical change may have resulted from the proliferation of such Internet-based social networks.[1]

navigation

- Main page
- Contents

Υπάρχουν κοινότητες στις δημοφιλείς Web εφαρμογές;

Web communities in more formal terms…

Communities on the Web are groups of individuals who share a common interest, together with the Web pages most popular amongst them … explicit or implicit ?

Web community is a set of Web-based objects (documents and users) that has its own logical and semantic structures, such that information retrieval and Web-data management is facilitated. e.g. a Web page set with clusters in it is a community; web pages in a set that are related to a given Web page also form a community;

Communities are groups of vertices which probably share common properties and/or play similar roles within the graph, e.g. groups of Web pages dealing with related topics

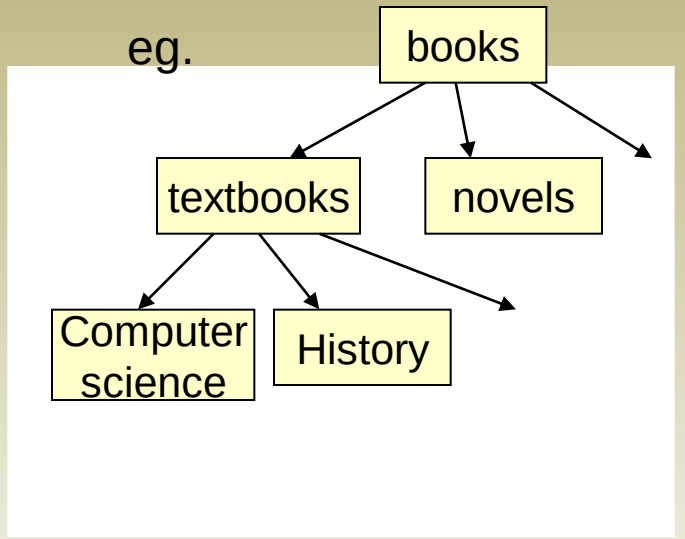
… two different types of communities

easily identified

- explicitly-defined communities

well-known group of web pages sharing a common interest e.g. Yahoo

eg.



(graph) analysis required

- implicitly-defined communities

non-obvious; hidden or unexpected;larger and outnumber the explicit ones; may appear as an emerging Web community for some specific topic or event;

eg. group of web pages for mediterranean cooking

focus on implicitly-defined communities

… in other terms

- *managerially coordinated communities:* they have a central authority or process (e.g. a creator) that governs the formation mechanism.
  - e.g. Google groups, LinkedIn, Drupal groups, Facebook, etc
- *self-organized communities:* they emerge from the interaction patterns between the members, i.e. highly related members are identified as a community.
  - e.g. Flickr clusters

communities context …

- typically … communities are defined with reference to some graph (network) which represents a set of entities / objects (nodes) and their relations (edges).

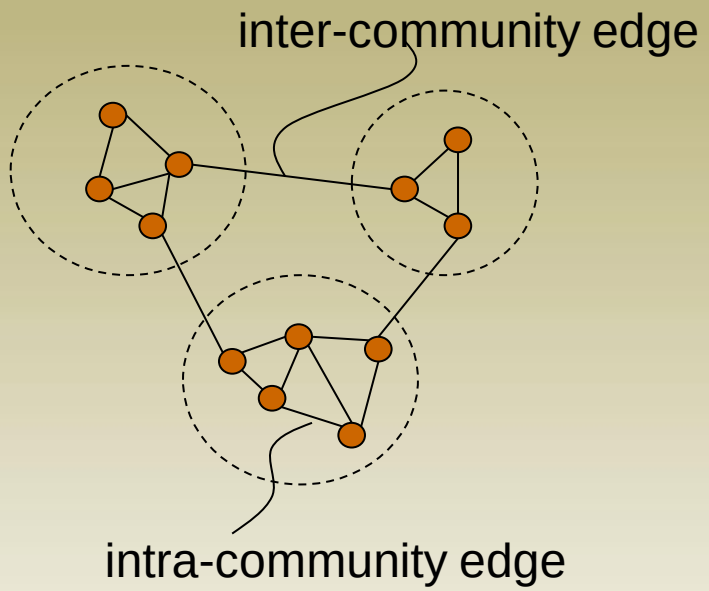
… even when there is no explicit graph, one can infer it, e.g.:

feature vectors → distances →threshold application → graph

- Given a graph, a community is loosely defined as a set of nodes that are more densely connected to each other than to the rest of the graph vertices.

A simple example …

- extremely profound community structure.
- key-concepts : within-community nodes, intra-community edges, inter-community edges.
- rarely appearing in real systems.

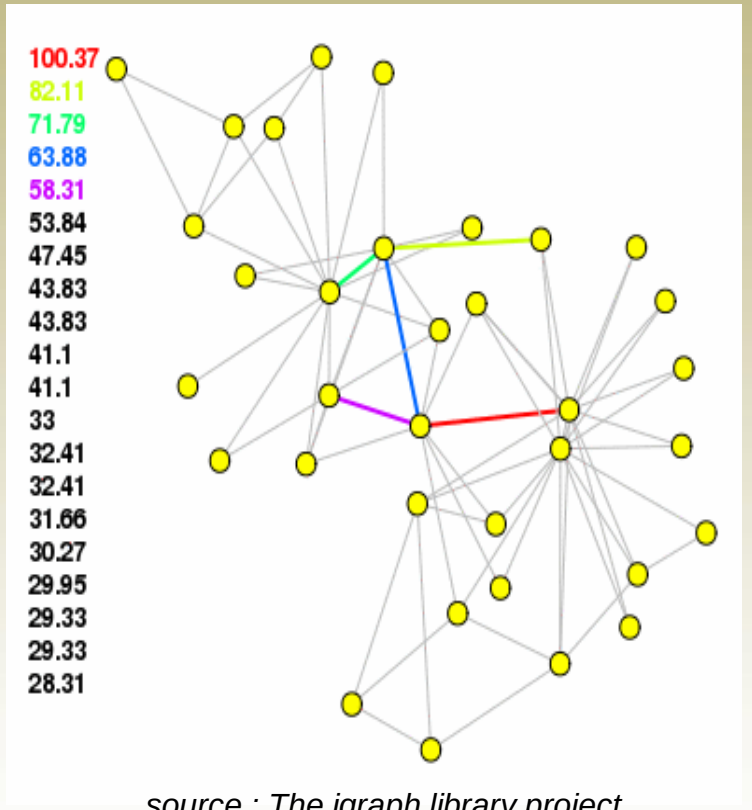


Definition of communities is heavily dependent on graph properties and subgraphs discovery

Edge centrality & betweenness

edge betweenness is the number of shortest paths between pairs of nodes that run along this edge. *If there is more than one shortest path between a pair of nodes, each path is assigned equal weight such that the total weight of all of the paths is equal to unity*.

- edges with the maximum score are considered as important for keeping the graph interconnected i.e. their removal results in unconnected clusters. High scoring edges are the "weak ties" or bottlenecks for traffic moving about the network

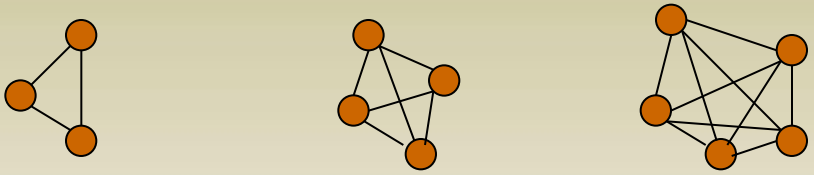


source : The igraph library project.

Subgraph structures (I)

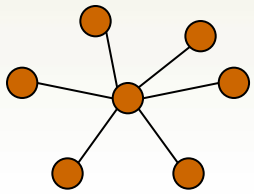
- *k-clique*: maximal subgraph each vertex of which is adjacent to all others. k-clique problem is in class NP, i.e. there is no solution to finding cliques in polynomial time, but if a clique is found in the graph, it can be verified in polynomial time.

k=3 (triangle) k=4 k=5



too strong for many purposes

- *N-clique*: maximal subgraph in which the maximum distance between two vertices is N, i.e. N stands for the length of the path allowed to make a connection to all other members (e.g. if path distance is 2 this corresponds to being "a friend of a friend.“)



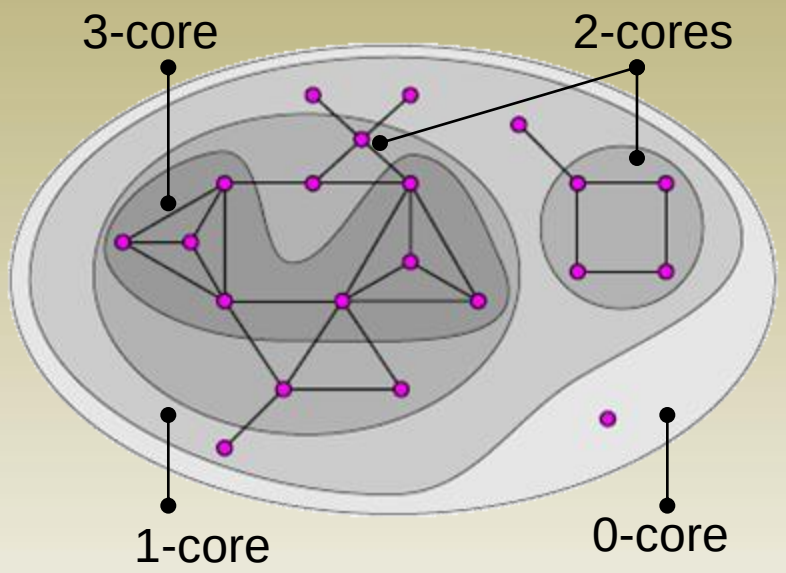
N=2 (star)

more helpful and general

Subgraph structures (II)

- k-core: maximal subgraph in which each vertex is connected to at least k other vertices. k-core is computed by pruning all the vertices (with their respective edges) with degree less than k. this operation can be useful to filter or to study some properties of the graphs

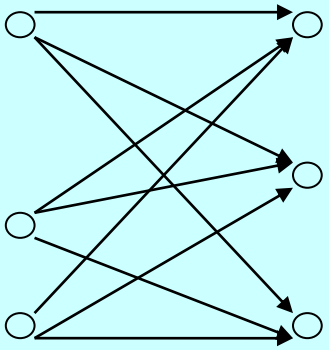
- k-plex: maximal subgraph in which each vertex is connected to all vertices except at most k of them (complementary to k-core)



k-core approach is more relaxed, by including vertices once they are connected to k members, regardless of how many other members they may not be connected to.

Specific graphs/subgraphs (II)

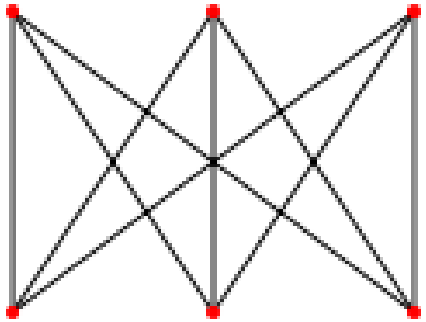
- **complete bipartite graph** or **biclique:** a special kind of bipartite graph where every vertex of the first set is connected to every vertex of the second set.



A (i=3, j=3) bipartite core

- **Bipartite core :** a complete bipartite subgraph with at least i nodes from U and at least j nodes from V

  i and j are tunable parameters



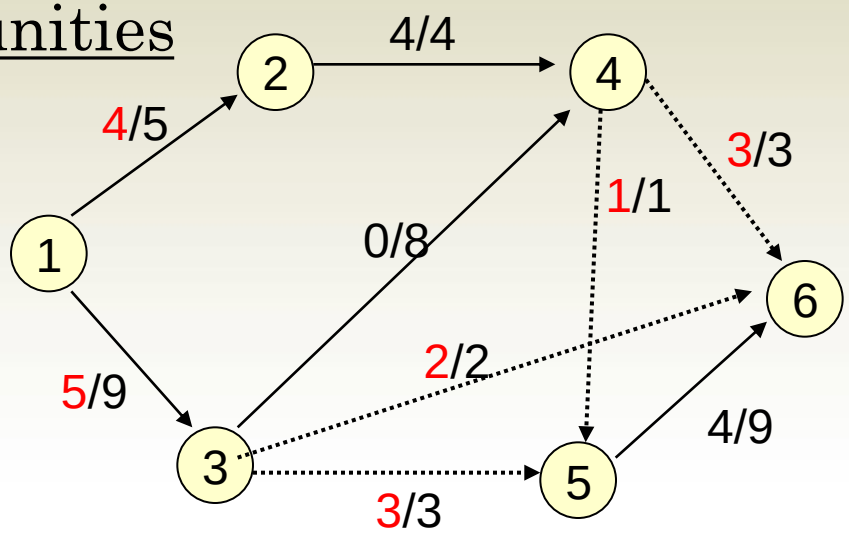
The notion of graph cut

- Given two vertices u and v of graph *G(V,E)*

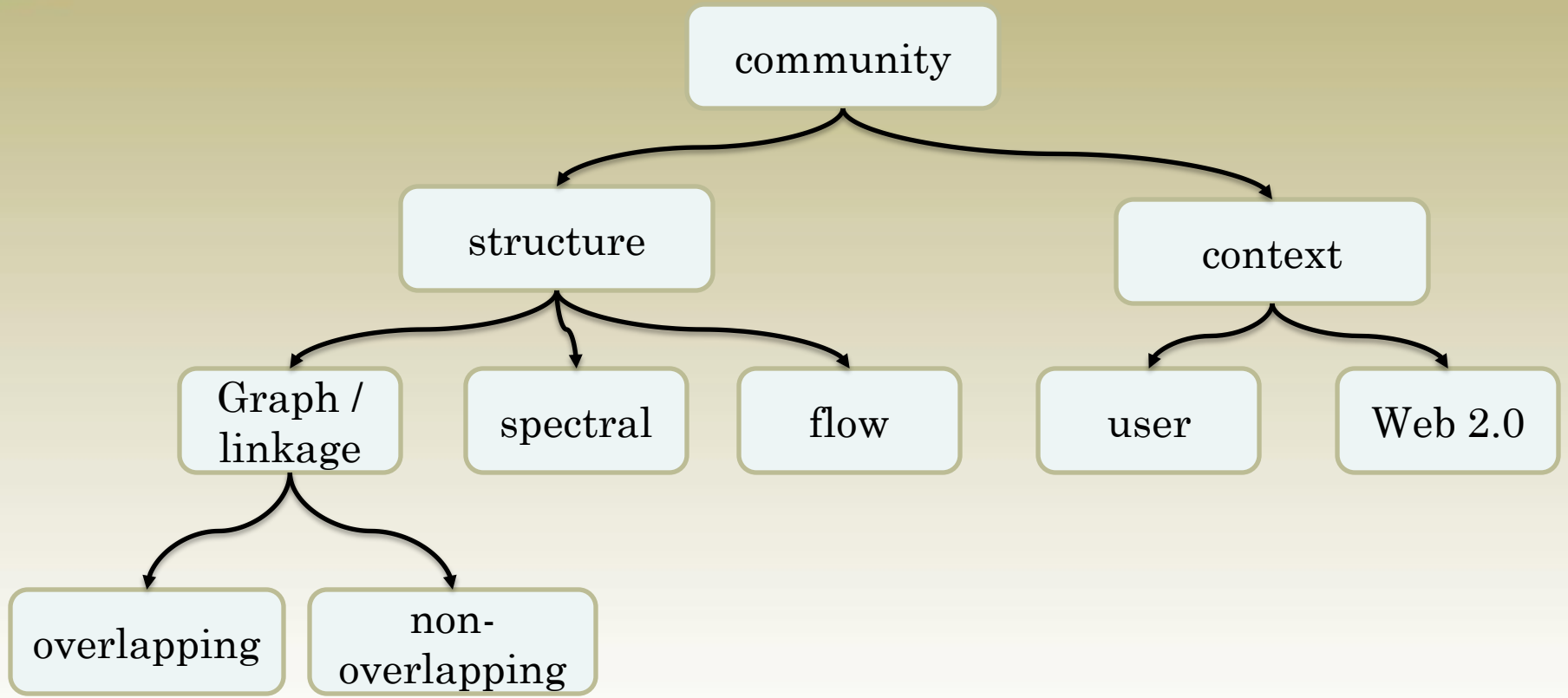
  a **cut** is a set of edges C ⊂ E such that the two vertices become separated on the induced graph *G' (V, E \ C)*
- denoting a source terminal as *s* and a sink terminal as *t*, a **cut** (*S*, *T*) of *G(V,E)* is a partition of *V* into *S* and *T* = *V* \ *S*, such that s ∈ *S* and t ∈ *T*
- ***Max-flow - min-cut theorem*** *originally* proposed by Ford and Fulkerson (1956) proves that maximum flow of the network is identical to minimum cut that separates *s* and *t*

The role of the flows

- **Definition** (*s*-*t* Maximum Flow)**:** given a directed graph, $G=(V,E)$, with edge capacities $c(u,v) \geq 0$, and two vertices, $s, t \in V$, find the maximum flow that can be routed from the source, *s*, to the sink, *t*
- **Intuition:** think of water pipes
- **Note:** maximum flow = minimum cut.
- Maximum flow yields communities



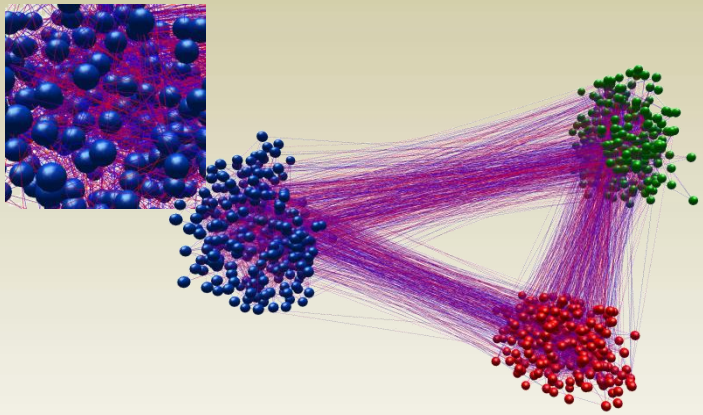
Ορισμοί κοινοτήτων: Κατηγοριοποίηση



Emphasis on linkage density

communities are subsets of vertices within which vertex-vertex connections are dense, but between which connections are less dense [Girvan02], [Newman04c] .

communities in the Web might represent pages on related topics.



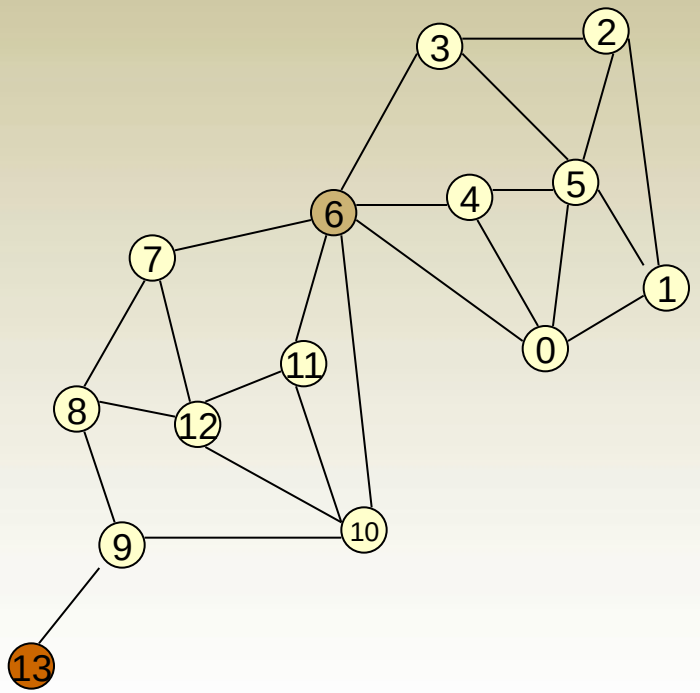
Communities are groups of vertices which probably share common properties and/or play similar roles within the graph. [Fortunato07a] .

communities may correspond to groups of pages of the World Wide Web dealing with related topics

Cores & communities

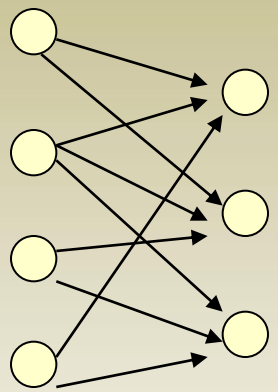
❖A bipartite core is the identity of a community and to extract all the communities is to enumerate all the bipartite cores on the web

- A web community is a (μ,ε)-core which is defined starting from a core vertex if its ε-neighborhood contains at least μ vertices
- A cluster consists of all vertices that belong to a core's neighborhood
  - hubs: isolated vertices connected to more than one clusters w.r.t. ε and μ.
  - outliers: isolated vertices connected to only one (μ,ε)-core.

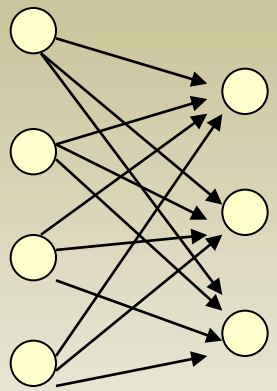


Web communities through bipartite graphs

- Dense bipartite graph

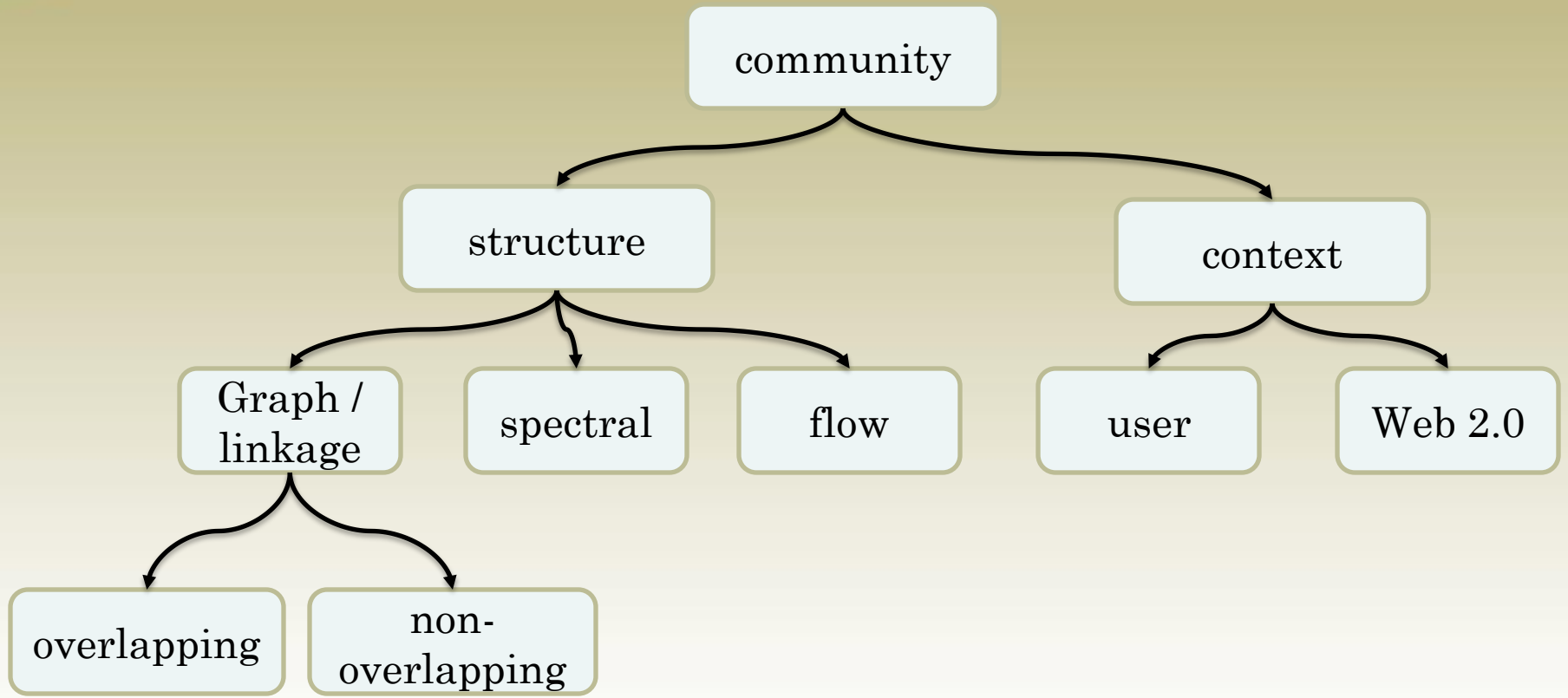


- Complete bipartite graph



- A Web community is a set of pages that form a dense bipartite graph
- A community is defined as a set of web pages whose hyperlinks form a complete bipartite graph ... since a complete bipartite graph abstraction extracts a small set of potential members to agree on some common interests

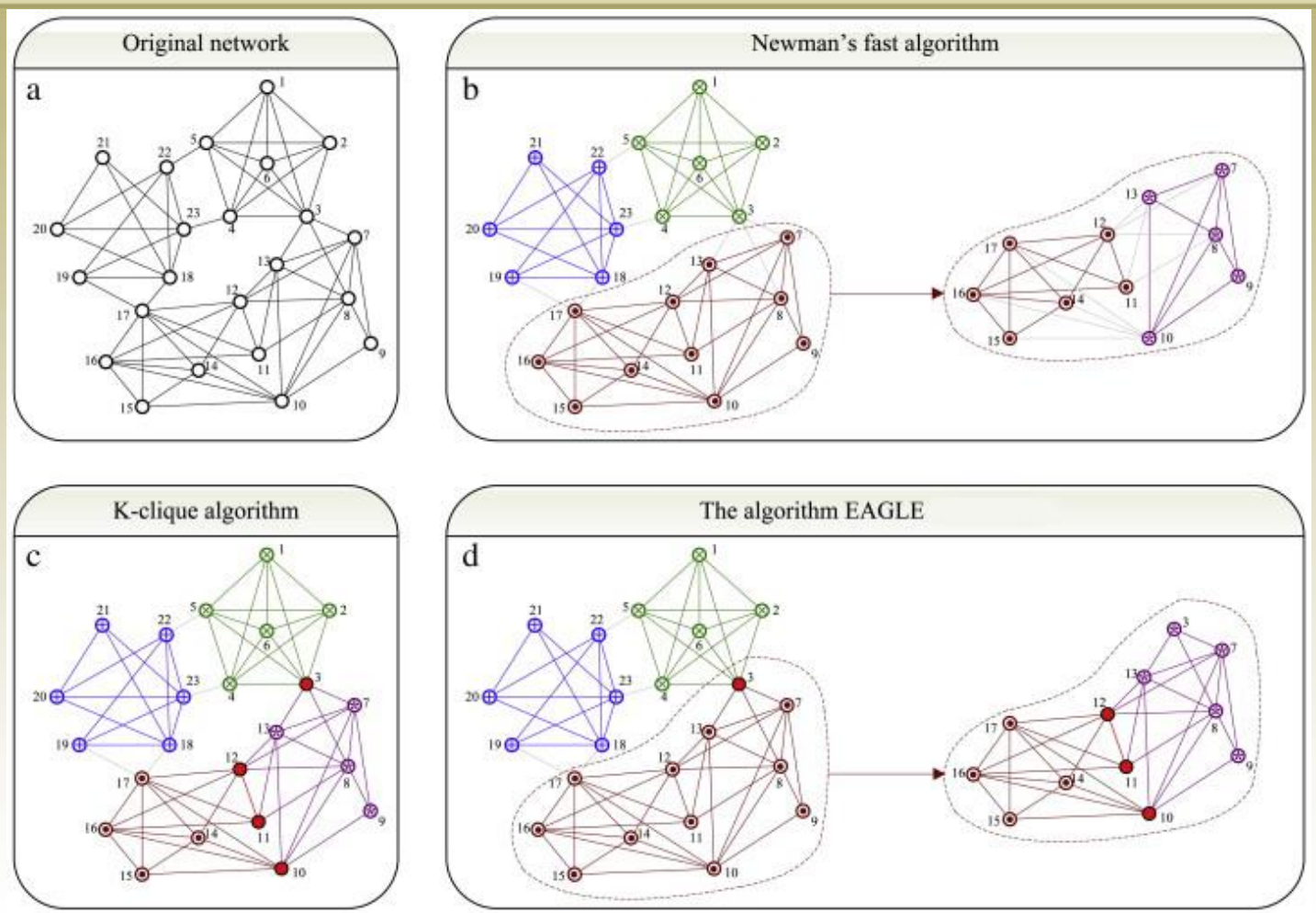
Ορισμοί κοινοτήτων: Κατηγοριοποίηση



Επικαλυπτόμενες κοινότητες

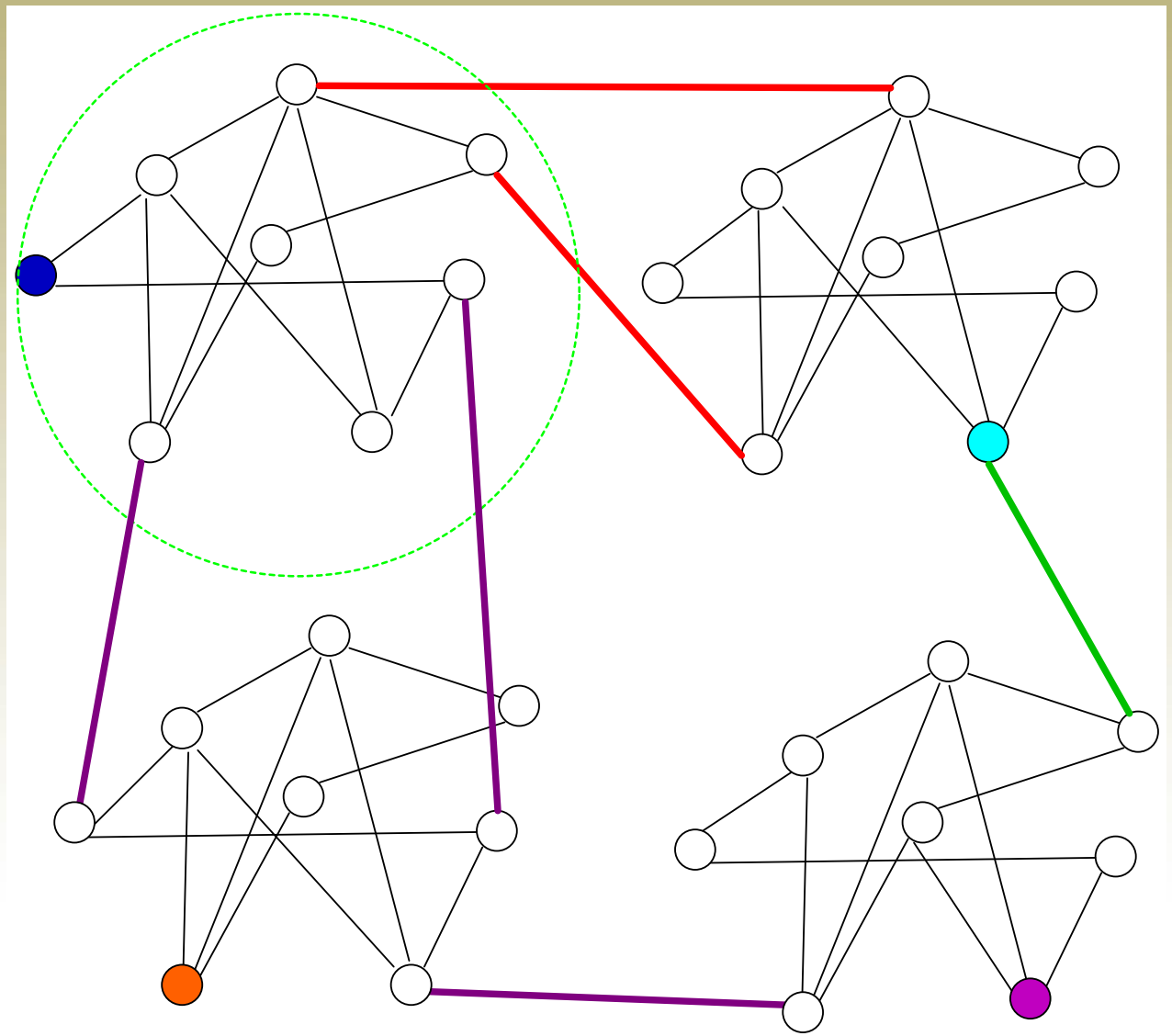
- majority of community definitions: driven by **crisp** clustering
- why not consider **non-crisp** clustering?
  - i.e., a vertex may belong to multiple communities
- it better fits human intuition
- still, does not require any text processing
- since communities are not disjoint, different sets of communities are solutions
- scalability to graphs with millions of edges is the **challenge**
- in many cases: necessary in order to produce reasonable results (see next slide)

Επικάλυψη σε κοινότητες



Different communities are rendered in different colors. Edges between communities are colored in light gray. Overlapping regions between communities are emphasized in red

Υπολογισμός της edge betwenness



Αλγόριθμος των Girvan-Newman (GN)

1. Υπολογισμός της betweenness για όλες τις ακμές του δικτύου

2. Αφαίρεση της ακμής με την μεγαλύτερη betweenness

3. Επανυπολογισμός της betweennesses για όλες τις ακμές, οι οποίες επηρεάζονται από την αφαίρεση

4. Επανάληψη από το Βήμα 2, μέχρι να εξαλειφθούν όλες οι ακμές

Ανάλυση του αλγορίθμου των GN

- To improve performance, the betweennesses can be calculated by using the fast algorithm of Newman , which calculates betweenness for all *m* edges in a graph of *n* vertices in time *O*(*mn*)
- Because this calculation has to be repeated once for the removal of each-edge, the entire algorithm runs *in worst-case time O* (*m^2 n*)
- However, after the removal of each edge, we only have to recalculate the betweennesses of those edges that were affected by the removal, which is at most only those in the same component as the removed edge. This means that running time may be better than worst-case for networks with strong community structure

Alternative improvements

- To try to reduce the running time of the algorithm further, one might be tempted to calculate the betweennesses of all edges only once and then remove them in order of decreasing betweenness
- However, It is found that this strategy does not work well, because if two communities are connected by more than one edge, then there is no guarantee that all of those edges will have high betweenness—we only know that at least one of them will

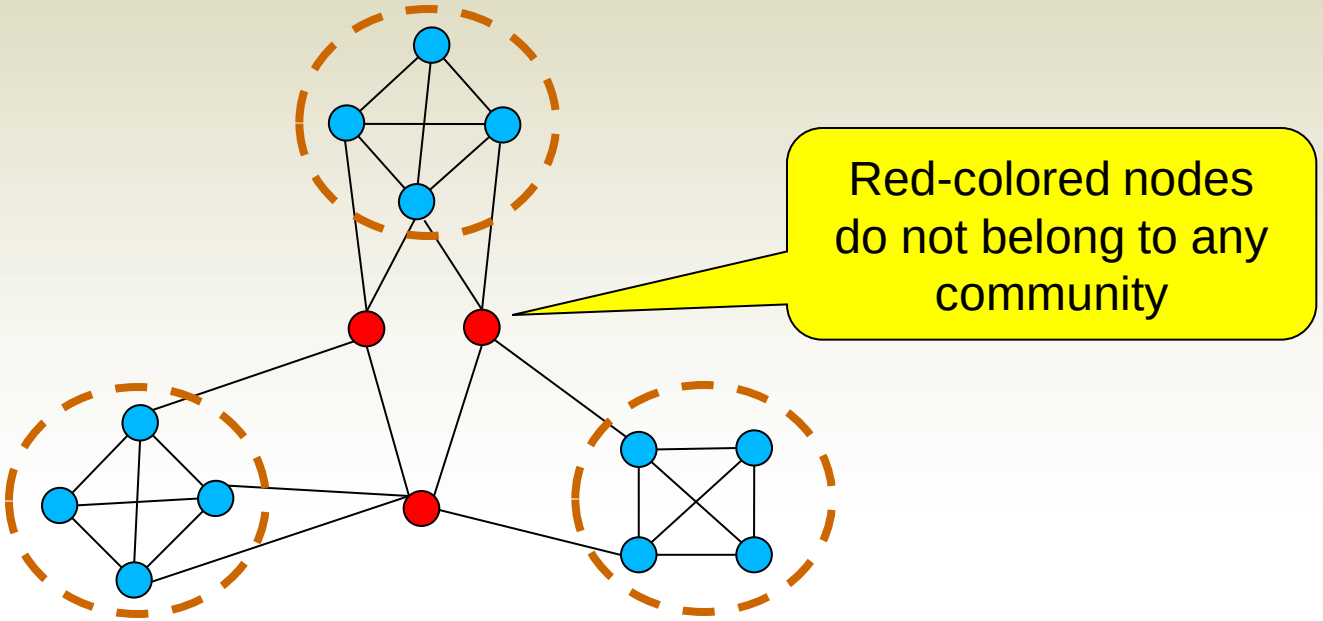
Ισχυρές κοινότητες

community in a strong sense

- each node in a strong community has more connections within (intra-) the community than with the rest (inter-) of the graph

$$k_i^{in}(V) > k_i^{out}, \forall i \in V$$

vertex-local focus

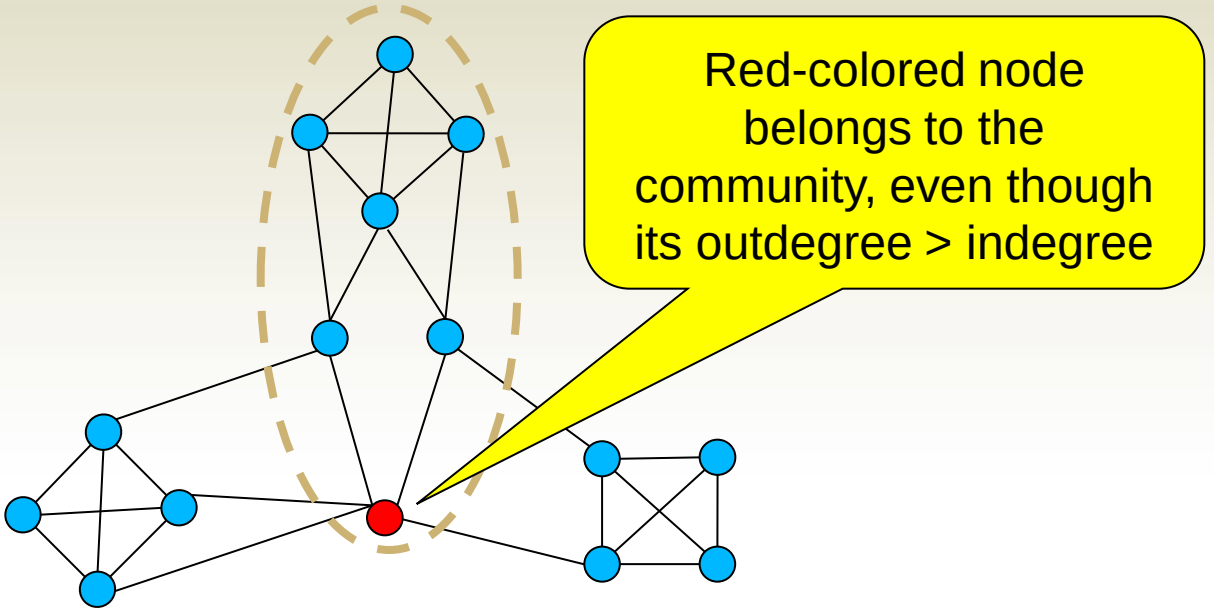


Ασθενείς κοινότητες

- community in a weak sense
  - in a weak community the sum of all degrees within V is larger than sum of all degrees toward the rest of the network

$$\sum_{i \in V} k_i^{in}(V) > \sum_{i \in V} k_i^{out}, \ \forall i \in V$$

community-global focus

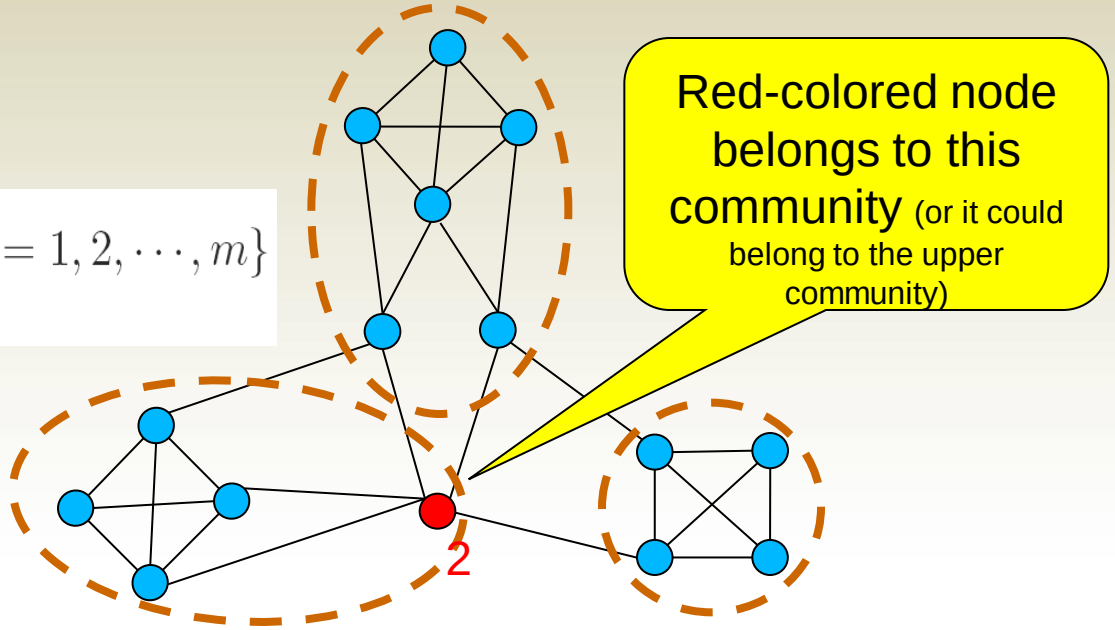


Ορισμός για “ασθενέστερο από τον ισχυρό” ορισμό

- Κοινότητα με πιο `ασθενή’ ορισμό
  - each node’s degree inside the “weaker than the strong” community should not be smaller than the node’s degree toward any other community
  - Εάν $V_1, V_2, \ldots, V_m$ είναι m communities του G, V_k, $k=1,2,\ldots,m$, πρέπει να ισχύει:

$$\bigcup_{k=1}^{k=m} V_k = G$$ και

$$\forall\, j \in V_k, \sum_{i \in V_k} A_{i,j} \geq max\{\sum_{i \in V_t} A_{i,j}, t = 1, 2, \cdots, m\}$$



Ασκήσεις

- **Άσκηση 1.**

  Στο παρακάτω δίκτυο για το οποίο δίνονται οι edge betweenness centralities, να βρείτε τις κοινότητες κατά Girvan-Newman, χωρίς να επανυπολογίζετε τις edge betweenness centralities μετά από κάθε αφαίρεση ακμής, και να κάνετε το δενδρόγραμμα.

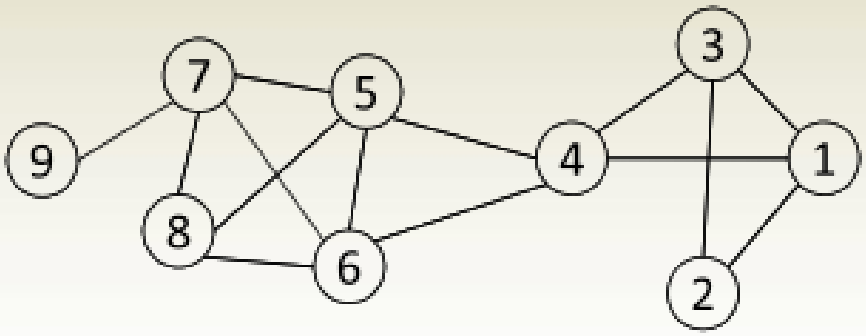


Table 3.3: Edge Betweenness

	1	2	3	4	5	6	7	8	9
1	0	4	1	9	0	0	0	0	0
2	4	0	4	0	0	0	0	0	0
3	1	4	0	9	0	0	0	0	0
4	9	0	9	0	10	10	0	0	0
5	0	0	0	10	0	1	6	3	0
6	0	0	0	10	1	0	6	3	0
7	0	0	0	0	6	6	0	2	8
8	0	0	0	0	3	3	2	0	0
9	0	0	0	0	0	0	8	0	0