Shaping Communities out of Triangles

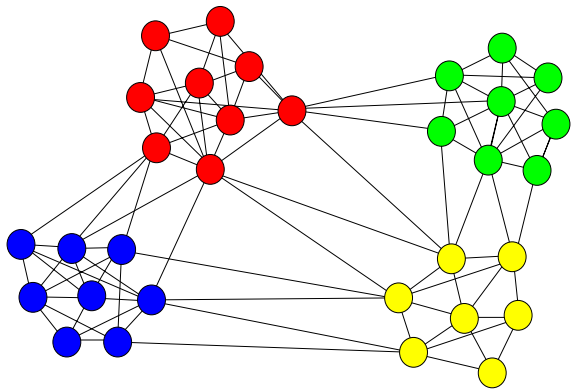
Arnau Prat-Pérez, David Dominguez-Sal, Josep M. Brunat and Josep-Lluis Larriba-Pey

DAMA-UPC, Universitat Politècnica de Catalunya

February 18, 2013

Introduction

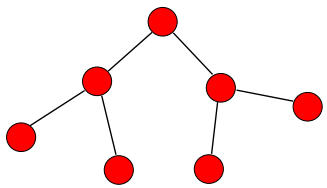
- A graph is a set of nodes connected by edges.
- A graph is used to represent interrelated data:
  - Social networks, Biology, The Web, etc...
- Communities are sets of densely connected nodes.



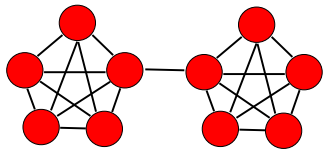
- Nodes in a community have similar characteristics.
- Useful for marketing purposes, new drug development, etc...

State of the Art Community Detection

- In the last years, a lot of metrics have been proposed ...
  - Modularity: Deviation from average density.
  - Conductance: Frontier vs. total edges.
- ... however, at certain circumstances, they fail.



Tree-like communities

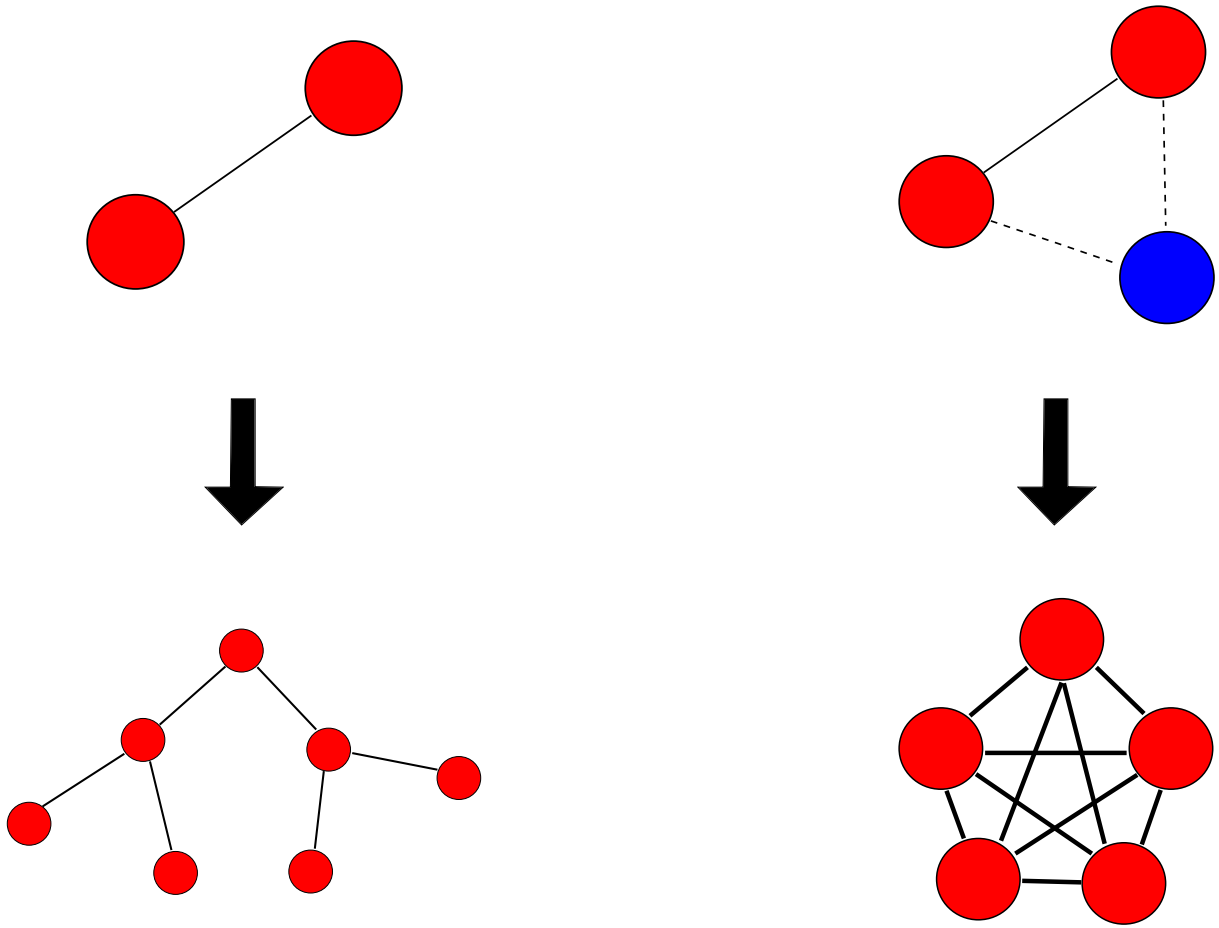


Chains of cliques

- **Reason:** they ignore the internal structure.

Contribution 1: A new metric based on triangles

- We propose Weight Community Clustering (WCC), a new metric based on triangles.

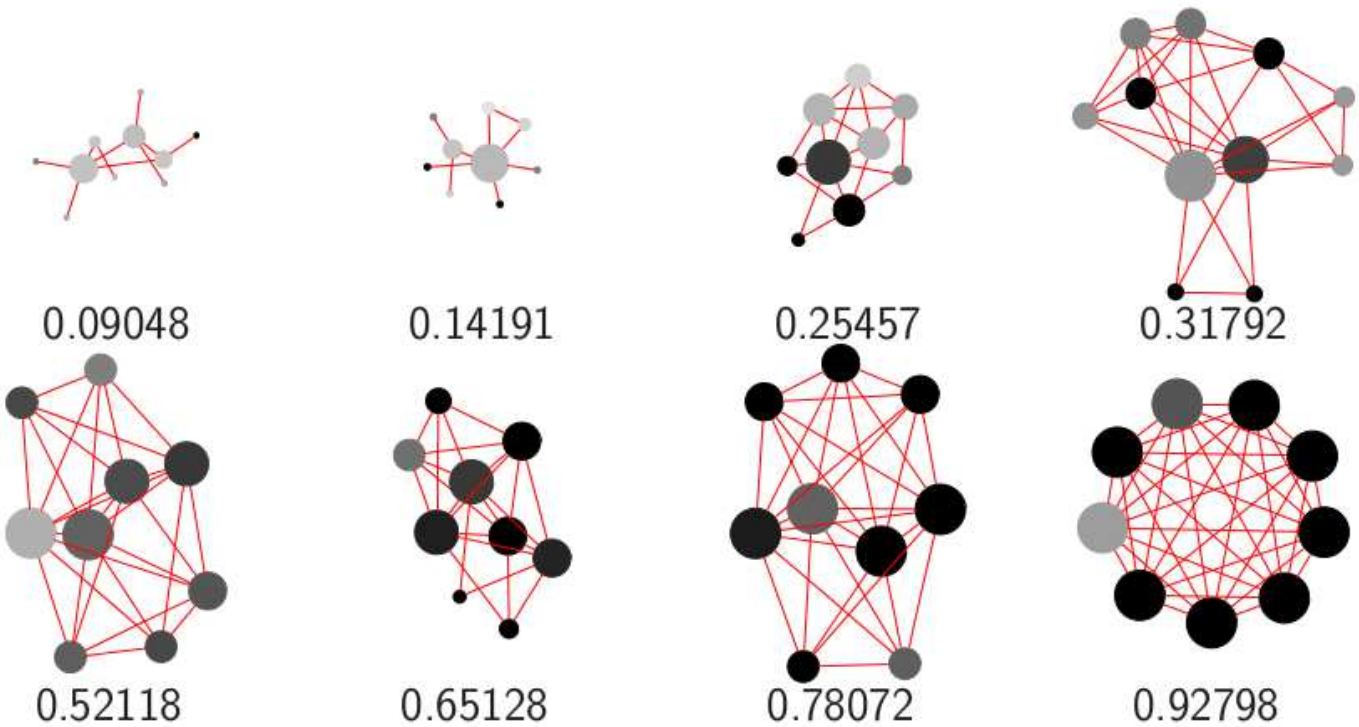


Contribution 2: formal guarantees

- We proof that WCC fulfills a basic set of properties:
  1. Communities have an internal structure.
  2. Linear community cohesion.
  3. Do not contain bridges.
  4. Avoid cut nodes.
- State of the art metrics **do not** formally prove their correctness.

Example

- Larger Size (+) => More connections inside the community
- Darker Colour (+) => Smaller edge cut.



Experimental results

- Good/Bad WCC $\rightarrow$ Good/Bad community.
- Good WCC tends to:
  - Large internal density of edges
  - Small diameter
  - Small cut ratio
  - Small bridge ratio
  - Large modularity
  - Small conductance
- We analyze the most relevant community algorithms in the state of the art by means of WCC.

Conclusions

- We propose a new metric based on triangles: WCC.
- We proof that WCC offers guarantees.
- Good/Bad WCC $\rightarrow$ Good/Bad community.
- Algorithm for WCC maximization.
- Developing overlapped generalization and algorithm
- Thank you and see you in the poster presentation session.