



GRANDATA[®]

Ciencia de Redes (Humanas y Sociales)

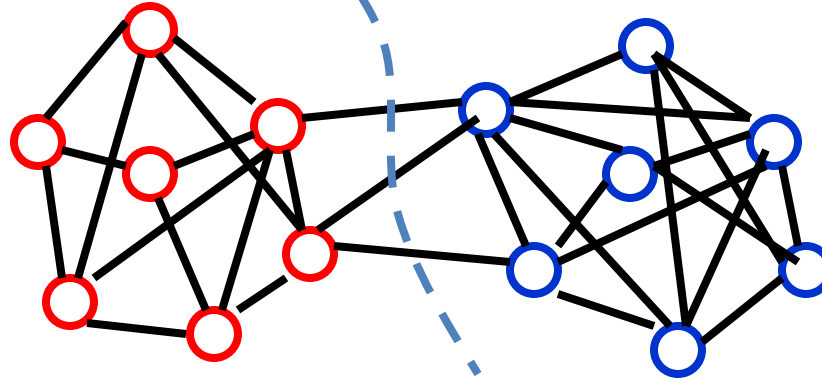
#3B

Carlos Sarraute

Instituto de Cálculo, Abril-Junio 2019

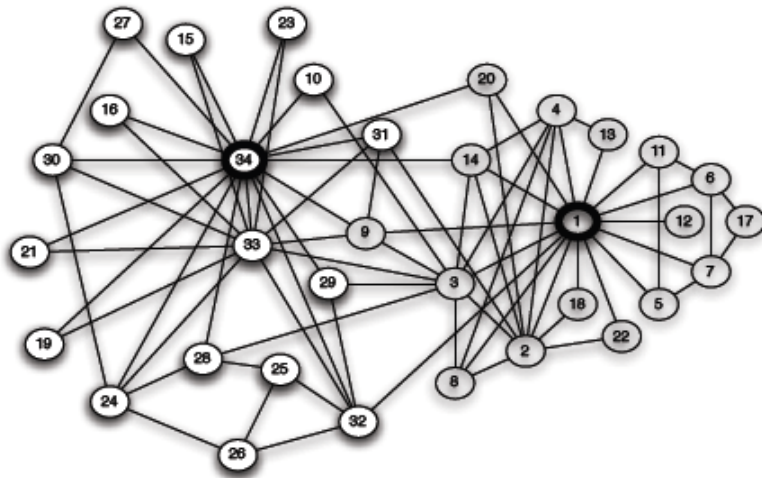
Enfoques para encontrar comunidades

- Las redes sociales y otras tienen una estructura en comunidades natural.
- Queremos descubrir esta estructura en lugar de imponer un cierto tamaño de comunidad o fijar el número de comunidades

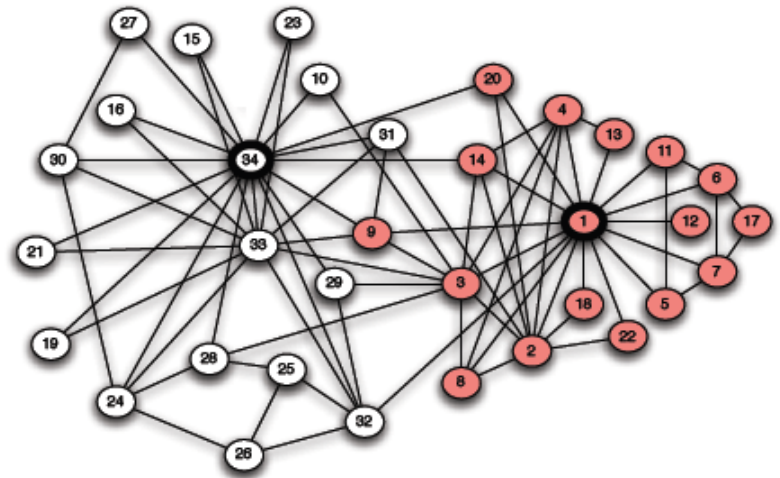


- Sin "mirar", ¿podemos descubrir la estructura de la comunidad de manera automatizada?

Zachary Karate Club



(a) *Karate club network*

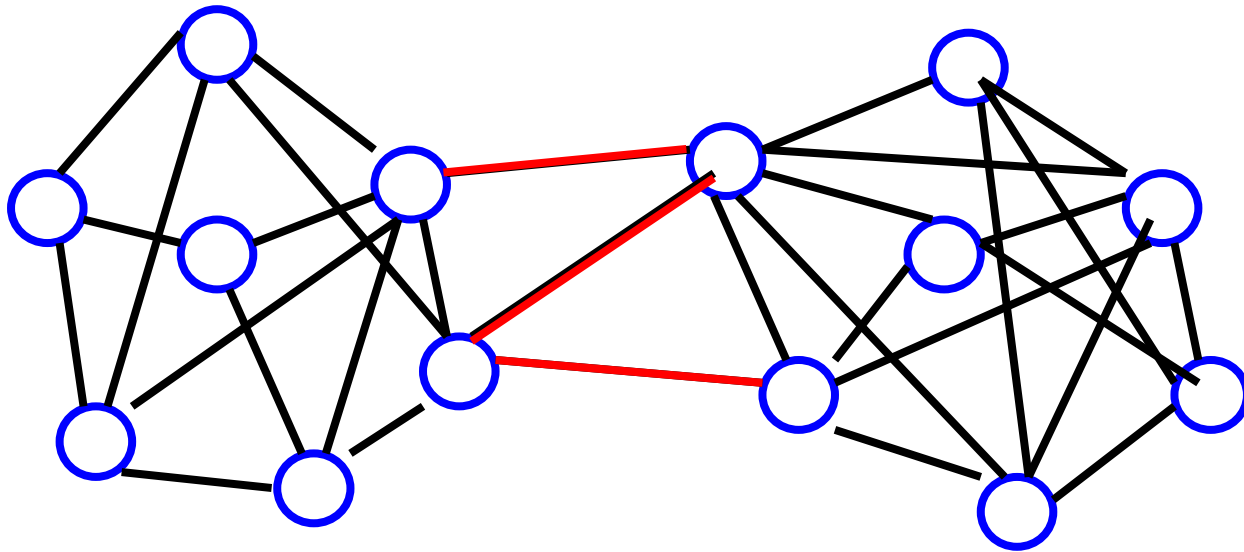


(b) *After a split into two clubs*

Betweenness clustering

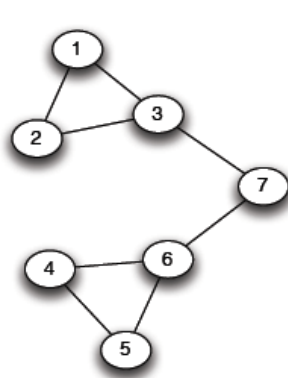
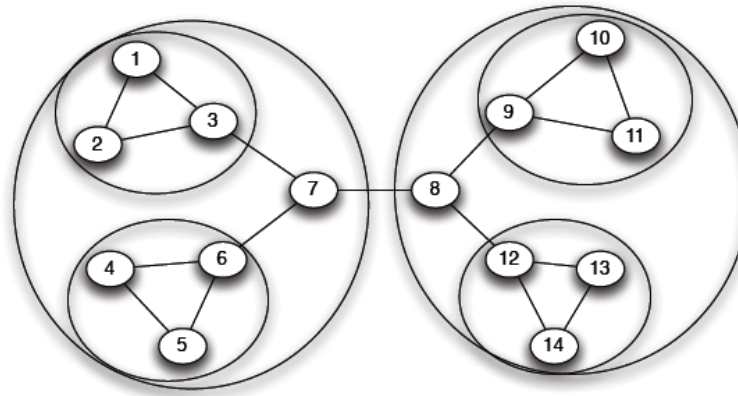
- Algorithm
 - compute the betweenness of all edges
 - while (betweenness of any edge > threshold):
 - remove edge with highest betweenness
 - recalculate betweenness
- Betweenness needs to be recalculated at each step
 - removal of an edge can impact the betweenness of another edge
 - very expensive: all pairs shortest path – $O(N^3)$
 - may need to repeat up to N times
 - does not scale to more than a few hundred nodes, even with the fastest algorithms

Betweenness clustering algorithm

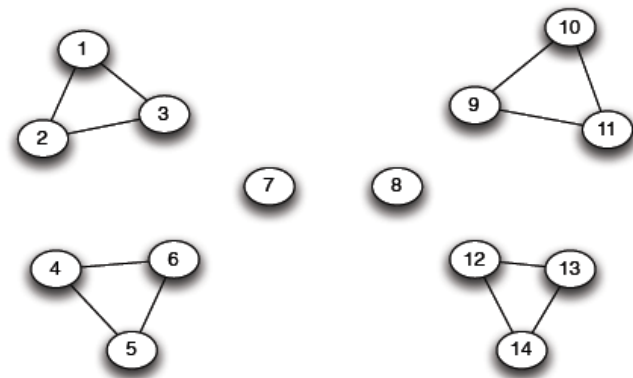


Betweenness clustering:

- successively remove edges of highest betweenness (the bridges, or local bridges), breaking up the network into separate components

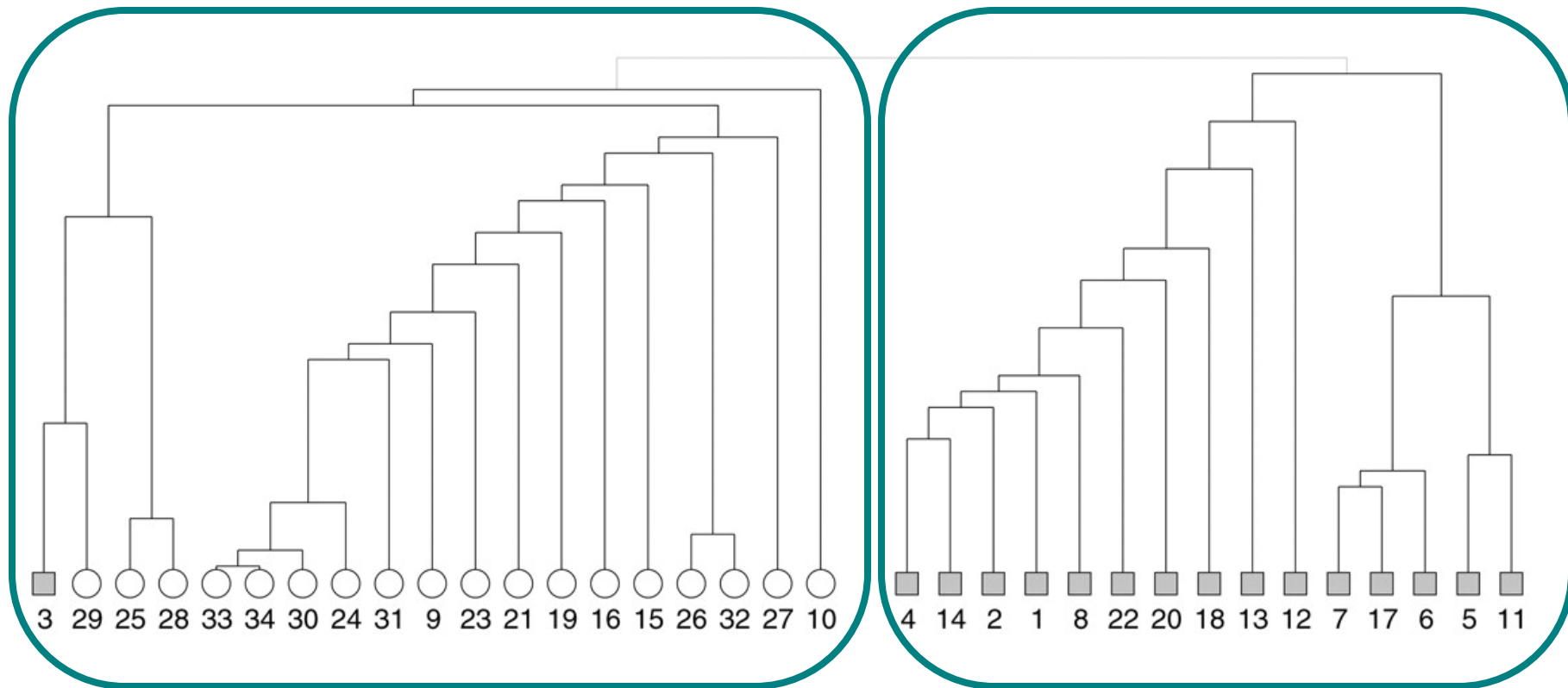


(a) Step 1



(b) Step 2

Algoritmo de Betweenness clustering sobre el dataset del club de karate



Modularidad

Algoritmo de Louvain

- <https://arxiv.org/pdf/0803.0476.pdf>

Fast unfolding of communities in large networks

Vincent D. Blondel^{1;a}, Jean-Loup Guillaume^{1,2;b}, Renaud Lambiotte^{1,3;c} and Etienne Lefebvre¹

¹Department of Mathematical Engineering, Université catholique de Louvain, 4 avenue Georges Lemaitre, B-1348 Louvain-la-Neuve, Belgium

² LIP6, Université Pierre et Marie Curie, 4 place Jussieu, 75005 Paris, France

³Department of Mathematical Sciences, Imperial College London, 53 Prince's Gate, London campus, SW72PG, UK

^avincent.blondel@uclouvain.be; ^bjean-loup.guillaume@lip6.fr; ^crenaud.lambiotte@imperial.ac.uk;



Tres hipotesis fundamentales

H1: Fundamental Hypothesis

- A network's community structure is uniquely encoded in its wiring diagram

H2: Connectedness and Density Hypothesis

- A community is a locally dense connected subgraph in a network

H3: Random Hypothesis

- Randomly wired networks lack an inherent community structure

Configuration Model

- Grafo con n nodos, m enlaces
- Cada nodo v tiene grado k_v
- Matriz de adyacencia A
 - $A_{vw} = 1$ si hay un enlace entre v y w
 - Grafo no dirigido $A_{vw} = A_{wv}$
- Configuration model:
 - Cortamos cada enlace en dos
 - Los reconectamos al azar
 - Incluso a sí mismo

Configuration Model

- La cantidad de cachos de enlace:

$$l_n = \sum_v k_v = 2m$$

- El nuevo grafo respeta los grados de los nodos
- Cuantos enlaces esperamos que haya entre dos nodos v, w ?

Configuration Model

- Cuantos enlaces esperamos que haya entre dos nodos v, w ?
 - Cantidad total de enlaces / cantidad de recableados posibles
- Recableados: $l_{n-1} \approx l_n$
- Cantidad enlaces esperados $= \frac{k_v k_w}{l_n} = \frac{k_v k_w}{2m}$

Modularidad Q

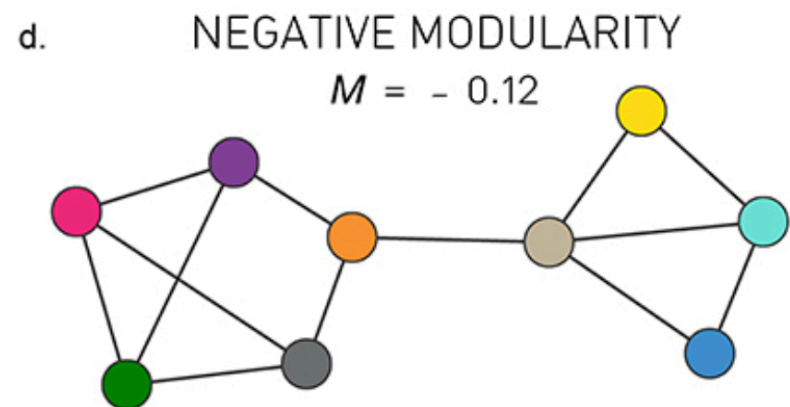
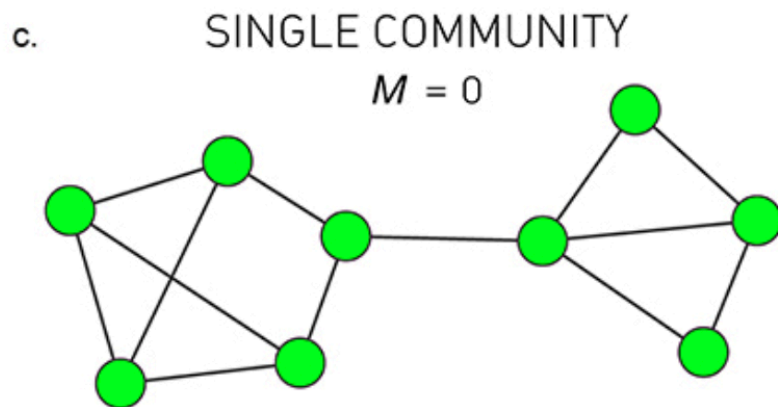
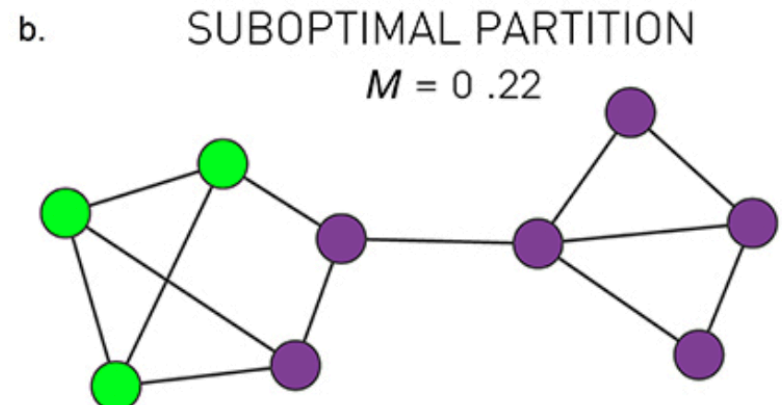
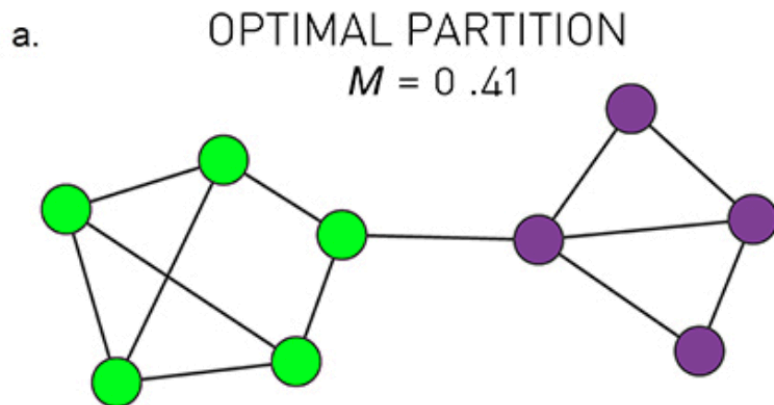
- La diferencia entre la realidad y el modelo random es

$$A_{vw} - \frac{k_v k_w}{2m}$$

- Sumando sobre todos los pares de nodos obtenemos la **modularidad Q**

$$Q = \frac{1}{(2m)} \sum_{vw} \left[A_{vw} - \frac{k_v k_w}{(2m)} \right] \delta(c_v, c_w)$$

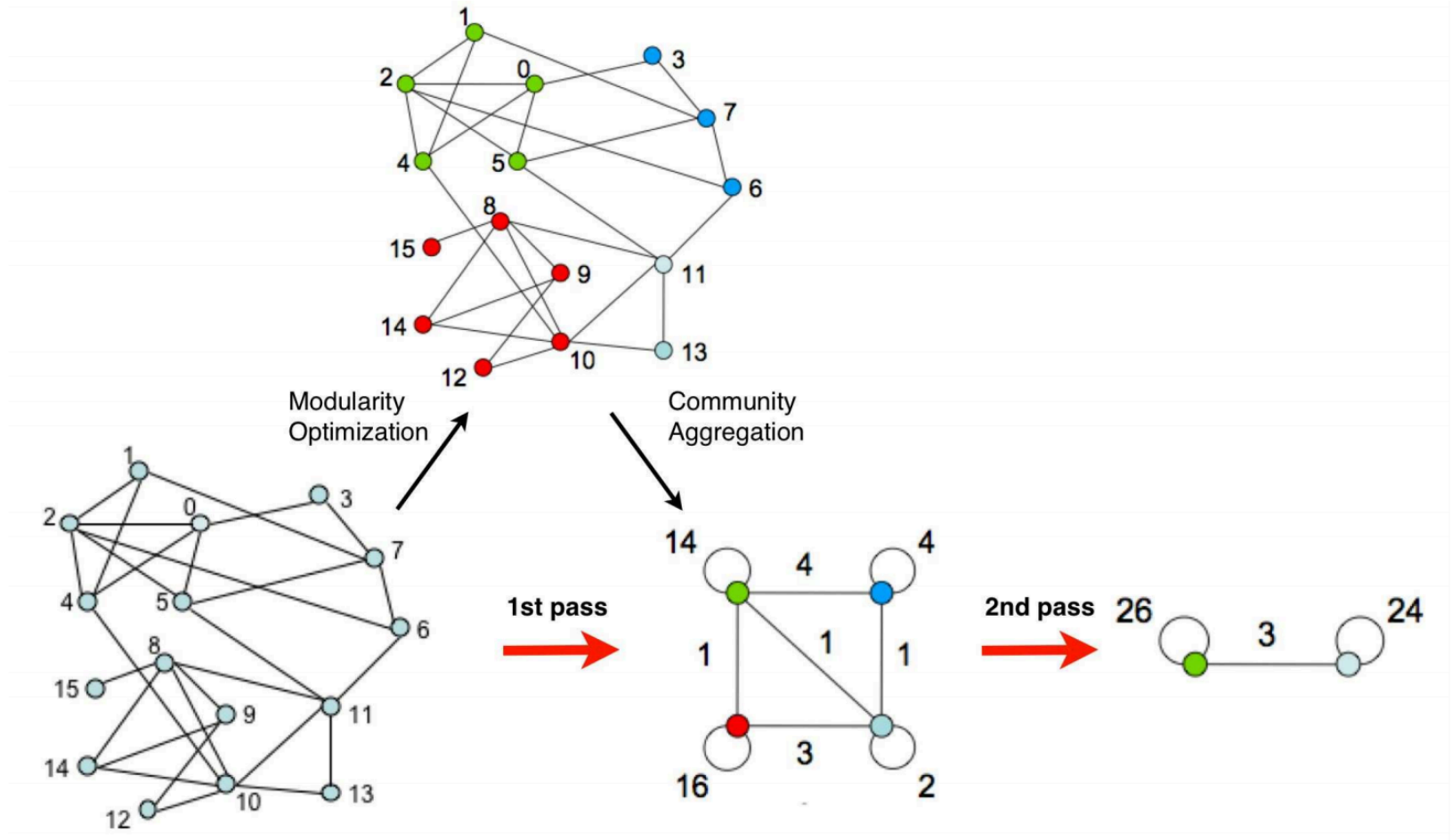
Modularidad



Método de Louvain

- Primero, cada nodo en la red se asigna a su propia comunidad.
- Luego, para cada nodo u , el cambio en la modularidad se calcula al eliminar u de su propia comunidad y moverlo a la comunidad de cada vecino v
- u se coloca en la comunidad que resultó en el mayor aumento de modularidad.
- Se repite hasta llegar a un máximo de la modularidad

Meta-comunidades



Red de celulares en Bélgica

