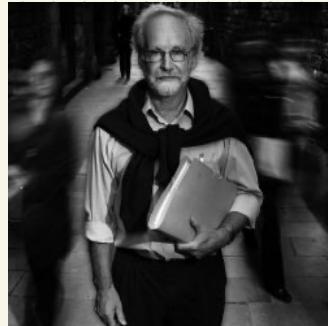


Modularity and Graph Expansion

Fiona Skerman

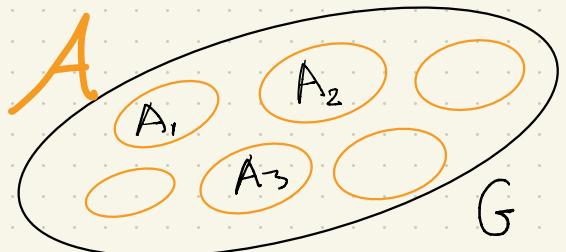
with Baptiste Louf + Colin McDiarmid

arXiv:2312.07521



Modularity 'meas. of how well a graph can be clustered'

NEWMAN + GIRVAN 2004.



graph G , m edges. $A = \{A_1, \dots, A_k\}$ vertex partition

score of partition A , $q_A(G) =$

modularity of G $q^*(G) = \max_A q_A(G)$

"high vals taken to indicate
more community structure"

Community Detection

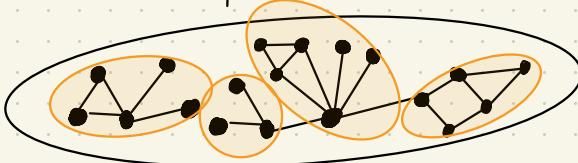
input graph $G = (V, E)$

vertices
nodes

edges
(weighted)

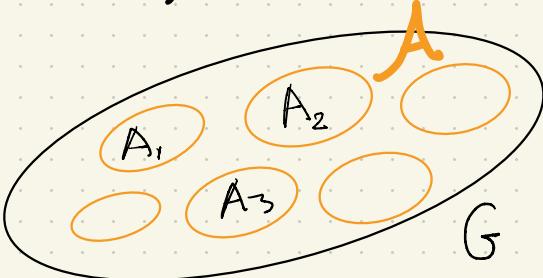
output vertex partition A
'community division'

G



- modularity score NP-hard to opt.
- Louvain ~ modularity based
 - ~ iteratively build a partition local choices - maximise mod.
- most popular methods use modularity

Modularity 'meas of how well a graph can be clustered'



graph G , m edges. $A = \{A_1, \dots, A_k\}$ vertex partition

score of partition A , $q_A(G) =$

modularity of G $q^*(G) = \max_A q_A(G)$

"high vals taken to indicate more community structure"

$$q_A(G) = \sum_{A \in A} \frac{e(A)}{m} - \frac{\text{vol}(A)^2}{(2m)^2} = \frac{1}{2m} \sum_{A \in A} \sum_{u, v \in A} \mathbb{1}_{[u \sim v]} - \frac{d_u d_v}{2m}$$

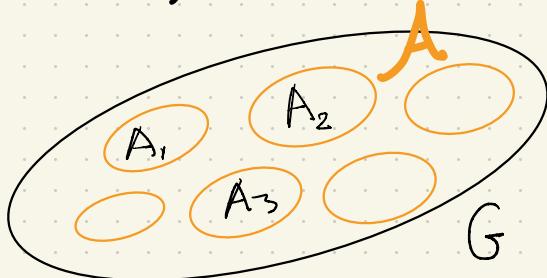
"edge contrib". "degree tax"

d_u = #edges incident to u

$\text{vol}(A)$ = #edges in set A

$$\text{vol}(A) = \sum_{u \in A} d_u$$

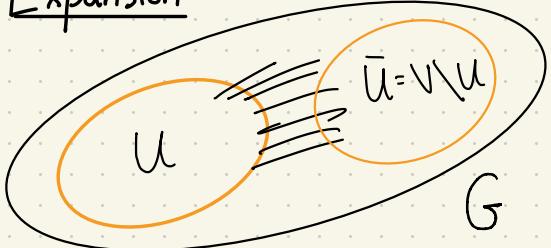
Modularity 'meas of how well a graph can be clustered'



$$q_A(G) = \sum_{A \in A} \frac{e(A)}{m} - \frac{\text{vol}(A)^2}{(2m)^2} = \frac{1}{2m} \sum_{A \in A} \sum_{u, v \in A} \mathbf{1}_{[u \sim v]} - \frac{d_u d_v}{2m}$$

"edge contrib" "degree tax"

Expansion



conductance

$$h_G = \min_{U \subseteq V(G)} \frac{e(U, \bar{U})}{\min \{ \text{vol}(U), \text{vol}(\bar{U}) \}}$$

$$\hat{h}_G = \min_{U \subseteq V(G)} \frac{e(U, \bar{U}) \text{vol}(G)}{\text{vol}(U) \text{vol}(\bar{U})}$$

graph G , m edges. $A = \{A_1, \dots, A_k\}$ vertex partition

score of partition A , $q_A(G) =$

modularity of G

$$q^*(G) = \max_A q_A(G)$$

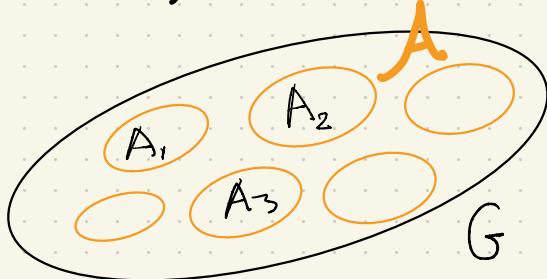
high vals taken to indicate
more community structure

d_u = #edges incident to u

$e(A)$ = #edges in set A

$$\text{vol}(A) = \sum_{u \in A} d_u$$

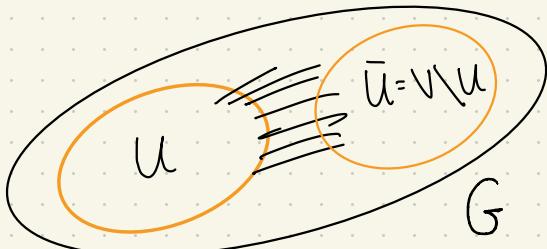
Modularity 'meas of how well a graph can be clustered'



$$q_A(G) = \sum_{A \in A} \frac{e(A)}{m} - \frac{\text{vol}(A)^2}{(2m)^2} = \frac{1}{2m} \sum_{A \in A} \sum_{u, v \in A} \mathbf{1}_{[u \sim v]} - \frac{\sum_{u \in A} d_u}{2m}$$

"edge contrib." "degree tax"

Expansion



graph G , m edges. $A = \{A_1, \dots, A_k\}$ vertex partition

modularity of G

$$q^*(G) = \max_A q_A(G)$$

"high vals taken to indicate more community structure"

$d_u = \# \text{edges incident to } u$

$e(A) = \# \text{edges in set } A$

$$\text{vol}(A) = \sum_{u \in A} d_u$$

conductance

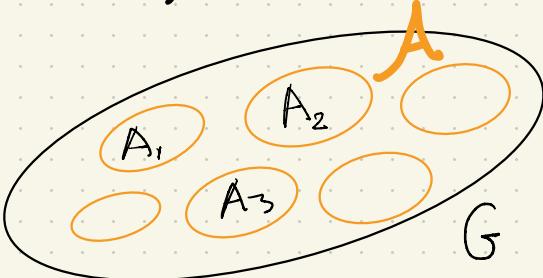
$$h_G = \min_{U \subseteq V(G)} \frac{e(U, \bar{U})}{\min\{\text{vol}(U), \text{vol}(\bar{U})\}}$$

$$\hat{h}_G = \min_{U \subseteq V(G)} \frac{e(U, \bar{U}) \text{vol}(G)}{\text{vol}(U) \text{vol}(\bar{U})}$$

Robustness $|q^*(G+!) - q^*(G)| < \frac{2}{e(G)}$

but $h_{G+!} = \hat{h}_{G+!} = 0$ disjoint!

Modularity 'meas of how well a graph can be clustered'



$$q_A(G) = \sum_{A \in A} \frac{e(A)}{m} - \frac{\text{vol}(A)^2}{(2m)^2} = \frac{1}{2m} \sum_{A \in A} \sum_{u, v \in A} \mathbb{1}_{[u \sim v]} - \frac{d_u d_v}{2m}$$

"edge contrib." "degree tax"

graph G , m edges. $A = \{A_1, \dots, A_k\}$ vertex partition

modularity of G

$$q^*(G) = \max_A q_A(G)$$

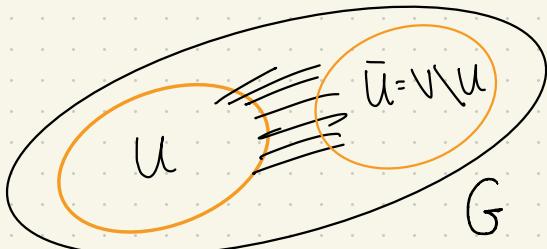
"high vals taken to indicate more community structure"

$d_u = \# \text{edges incident to } u$

$e(A) = \# \text{edges in set } A$

$$\text{vol}(A) = \sum_{u \in A} d_u$$

Expansion of Subgraphs



conductance

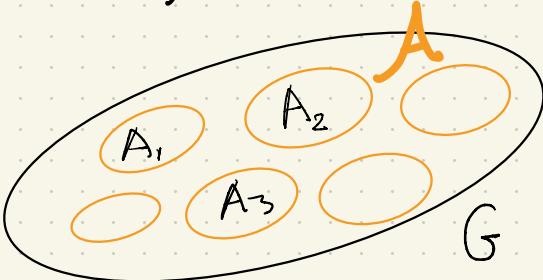
$$h_G = \min_{U \subseteq V(G)} \frac{e(U, \bar{U})}{\min\{\text{vol}(U), \text{vol}(\bar{U})\}}$$

$$\hat{h}_G = \min_{U \subseteq V(G)} \frac{e(U, \bar{U}) \text{vol}(G)}{\text{vol}(U) \text{vol}(\bar{U})}$$

Robustness $|q^*(G+!) - q^*(G)| < \frac{2}{e(G)}$

but $h_{G+!} = \hat{h}_{G+!} = 0$ disjoint!

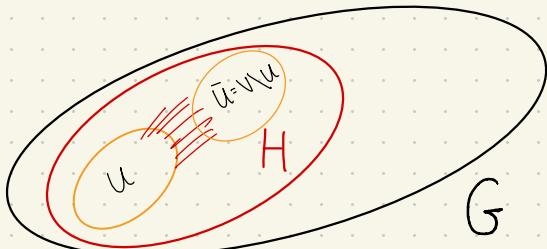
Modularity 'meas of how well a graph can be clustered'



$$q_A(G) = \sum_{A \in A} \frac{e(A)}{m} - \frac{\text{vol}(A)^2}{(2m)^2} = \frac{1}{2m} \sum_{A \in A} \sum_{u, v \in A} \mathbf{1}_{[u \sim v]} - \frac{d_u d_v}{2m}$$

"edge contrib." "degree tax"

Expansion of Subgraphs



$$\alpha = \frac{e(H)}{e(G)}$$

graph G , m edges. $A = \{A_1, \dots, A_k\}$ vertex partition

modularity of G

$$q^*(G) = \max_A q_A(G)$$

"high vals taken to indicate more community structure"

$d_u = \# \text{edges incident to } u$

$e(A) = \# \text{edges in set } A$

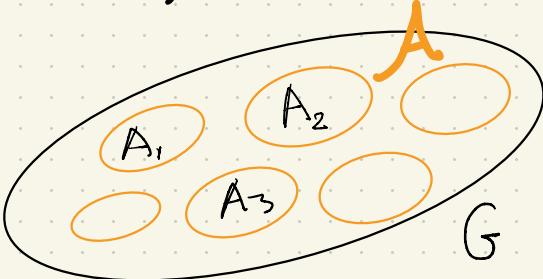
$$\text{vol}(A) = \sum_{u \in A} d_u$$

conductance

$$h_H = \min_{U \subseteq V(H)} \frac{e(U, \bar{U})}{\min \{ \text{vol}(U), \text{vol}(\bar{U}) \}}$$

$$\hat{h}_H = \min_{U \subseteq V(H)} \frac{e(U, \bar{U}) \text{vol}(H)}{\text{vol}(U) \text{vol}(\bar{U})}$$

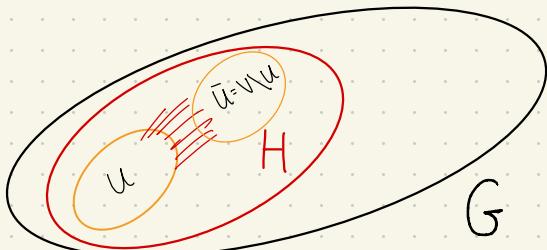
Modularity 'meas of how well a graph can be clustered'



$$q_A(G) = \sum_{A \in A} \frac{e(A)}{m} - \frac{\text{vol}(A)^2}{(2m)^2} = \frac{1}{2m} \sum_{A \in A} \sum_{u, v \in A} \mathbb{1}_{[u \sim v]} - \frac{d_u \cdot d_v}{2m}$$

"edge contrib." "degree tax"

Expansion of Subgraphs



$$\alpha = \frac{e(H)}{e(G)}$$

graph G , m edges. $A = \{A_1, \dots, A_k\}$ vertex partition

modularity of G

$$q^*(G) = \max_A q_A(G)$$

"high vals taken to indicate more community structure"

$d_u = \# \text{edges incident to } u$

$e(A) = \# \text{edges in set } A$

$$\text{vol}(A) = \sum_{u \in A} d_u$$

conductance

$$h_H = \min_{U \subseteq V(H)} \frac{e(U, \bar{U})}{\min\{\text{vol}(U), \text{vol}(\bar{U})\}}$$

$$\hat{h}_H = \min_{U \subseteq V(H)} \frac{e(U, \bar{U}) \text{vol}(H)}{\text{vol}(U) \text{vol}(\bar{U})}$$

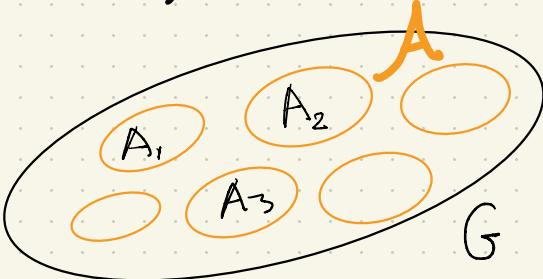
Thm (informal)

G has $q^*(G) \sim 1$

\Leftrightarrow

G has no large expander subgraphs

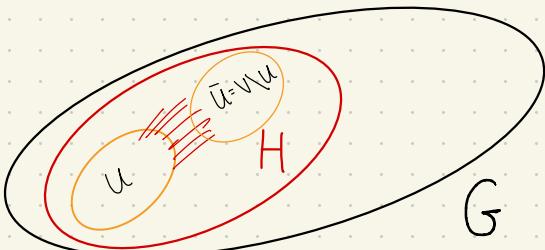
Modularity 'meas of how well a graph can be clustered'



$$q_A(G) = \sum_{A \in A} \frac{e(A)}{m} - \frac{\text{vol}(A)^2}{(2m)^2} = \frac{1}{2m} \sum_{A \in A} \sum_{u, v \in A} \mathbb{1}_{[u \sim v]} - \frac{d_u \cdot d_v}{2m}$$

"edge contrib." "degree tax"

Expansion of Subgraphs



$$\alpha = \frac{e(H)}{e(G)}$$

graph G , m edges. $A = \{A_1, \dots, A_k\}$ vertex partition

modularity of G

$$q^*(G) = \max_A q_A(G)$$

"high vals taken to indicate more community structure"

$d_u = \# \text{edges incident to } u$

$e(A) = \# \text{edges in set } A$

$$\text{vol}(A) = \sum_{u \in A} d_u$$

conductance

$$h_H = \min_{U \subseteq V(H)} \frac{e(U, \bar{U})}{\min\{\text{vol}(U), \text{vol}(\bar{U})\}}$$

$$\hat{h}_H = \min_{U \subseteq V(H)} \frac{e(U, \bar{U}) \text{vol}(H)}{\text{vol}(U) \text{vol}(\bar{U})}$$

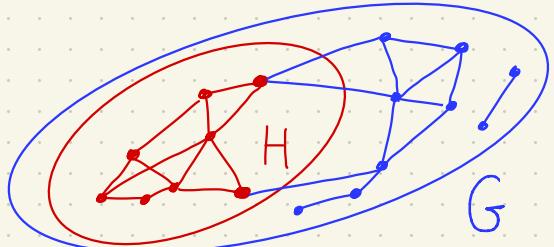
Thm (informal)

G has $q^*(G) \sim 1$

\Leftrightarrow

G has no large expander subgraphs

Modularity + Expansion of Subgraphs



$$\alpha = \frac{e(H)}{e(G)}$$

$$q_A(G) = \sum_{A \in A} \frac{e(A)}{m} - \frac{\text{vol}(A)^2}{(2m)^2}$$

"edge contrib." "degree tax"

$$q^*(G) = \max_A q_A(G).$$

$$h_H = \min_{U \subseteq V(H)} \frac{e(U, \bar{U})}{\min\{\text{vol}(U), \text{vol}(\bar{U})\}}$$

d_u = # edges incident to u

$e(A)$ = # edges in set A

$$\text{vol}(A) = \sum_{u \in A} d_u$$

Thm $\forall 0 < \alpha < 1, \forall \varepsilon > 0,$

(a) G has subgraph H , $\frac{e(H)}{e(G)} > \alpha$, $h_H \geq \alpha$

$$\Rightarrow q^*(G) \leq 1 - \alpha^2$$

(b) $\exists \delta > 0: q^*(G) \leq 1 - f(\alpha) - \varepsilon$

$\Rightarrow G$ has induced subgraph H ,

$$\frac{e(H)}{e(G)} > \alpha, \quad h_H \geq \delta.$$

$$f(\alpha) := \max \left\{ \sum x_i^2 : 0 \leq x_i \leq \alpha, \sum x_i = 1 \right\} \quad \text{NB. } f(\alpha) \approx \alpha \text{ for small } \alpha.$$

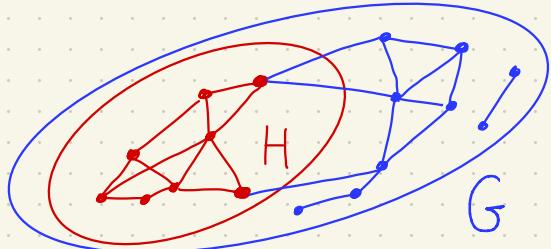
Thm (informal)

G has $q^*(G) \sim 1$

\Leftrightarrow

G has no large
expander subgraphs

Modularity + Expansion of Subgraphs



$$\alpha = \frac{e(H)}{e(G)}$$

Thm $\forall 0 < \alpha < 1, \forall \varepsilon > 0,$

(a) G has subgraph H , $\frac{e(H)}{e(G)} > \alpha$, $h_H \geq \alpha$

$$\Rightarrow q^*(G) \leq 1 - \alpha^2$$

(b) $\exists \delta > 0: q^*(G) \leq 1 - f(\alpha) - \varepsilon$

$\Rightarrow G$ has induced subgraph H ,

$$\frac{e(H)}{e(G)} > \alpha, \quad h_H \geq \delta.$$

$$f(\alpha) := \max \left\{ \sum x_i^2 : 0 \leq x_i \leq \alpha, \sum x_i = 1 \right\}$$

NB. $f(\alpha) \approx \alpha$ for small α .

$$q_A(G) = \sum_{A \in A} \frac{e(A)}{m} - \frac{\text{vol}(A)^2}{(2m)^2}$$

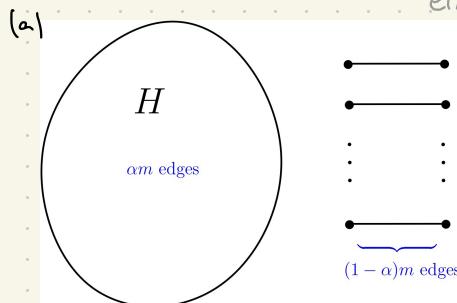
"edge contrib." "degree tax"

$$q^*(G) = \max_A q_A(G).$$

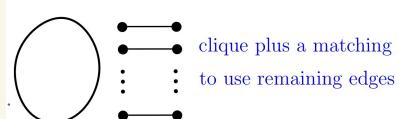
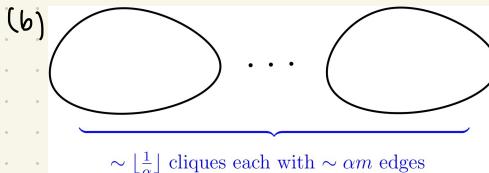
$$h_H = \min_{U \subseteq V(H)} \frac{e(U, \bar{U})}{\min \{ \text{vol}(U), \text{vol}(\bar{U}) \}}$$

$d_u = \# \text{edges incident to } u$

$e(A) = \# \text{edges in set } A$



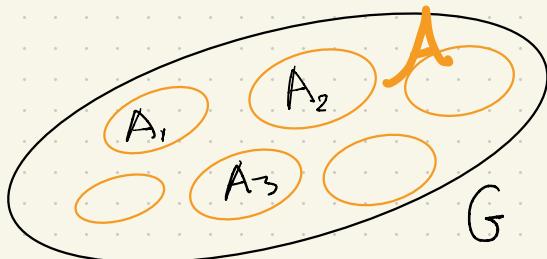
$$q^*(G_H) \geq 1 - \alpha^2 - \varepsilon$$



$$q^*(G_\alpha) \geq 1 - f(\alpha) + \varepsilon$$

any H , with $\frac{e(H)}{e(G)} > \alpha$
is disjoint.

Modularity 'meas of how well a graph can be clustered'



graph G , m edges. $A = \{A_1, \dots, A_k\}$ vertex partition

score of partition A , $q_A(G) =$

modularity of G

$$q^*(G) = \max_A q_A(G)$$

high vals taken to indicate
more community structure"

$$q_A(G) = \sum_{A \in A} \frac{e(A)}{m} - \frac{\text{vol}(A)^2}{(2m)^2} = \frac{1}{2m} \sum_{A \in A} \sum_{u, v \in A} \mathbf{1}_{[u \sim v]} - \frac{d_u d_v}{2m}$$

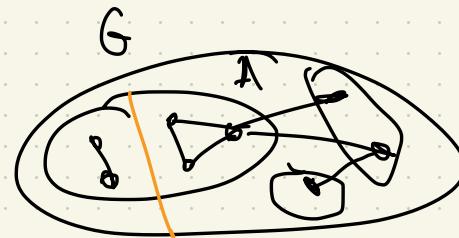
"edge contrib" "degree tax"

$d_u = \# \text{edges incident to } u$

$\text{vol}(A) = \# \text{edges in set } A$

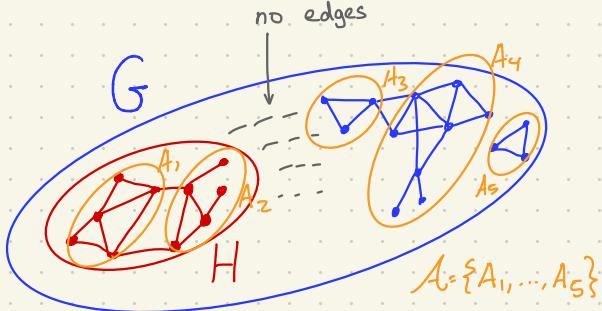
$$\text{vol}(A) = \sum_{u \in A} d_u$$

- A optimal partition of G i.e. $q_A(G) = q^*(G)$
- $\forall A \in A$ $G[A]$ conn. (+ isolated vert)



Modularity: Resolution Limit

$\text{OPT}(G)$: set of partitions achieving maximal modularity



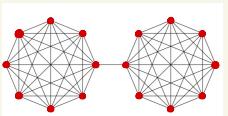
FORTUNATO + BARTHÉLEMY 2008.

Res Limit

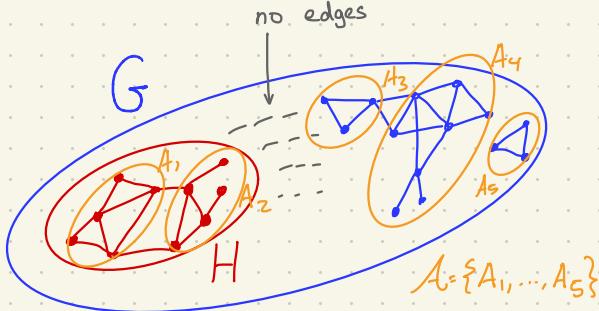
H is a connected component in graph G .

$$\cdot \forall H : e(H) < \sqrt{2 e(G)} \Rightarrow H \text{ not split}$$

$$\cdot H \text{ 'dumbbell graph'} \quad e(H) > \sqrt{2 e(G)} \Rightarrow H \text{ split}$$



Modularity: Resolution Limit



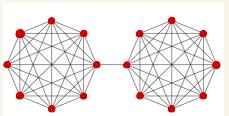
FORTUNATO + BARTHÉLEMY 2008.

Res Limit

H is a connected component in graph G .

$$\cdot \forall H : e(H) < \sqrt{2 e(G)} \Rightarrow H \text{ not split}$$

$$\cdot H \text{ 'dumbbell graph'} \quad e(H) > \sqrt{2 e(G)} \Rightarrow H \text{ split}$$



$\text{OPT}(G)$: set of partitions achieving maximal modularity

Thm H is a connected component in graph G . Let

$$\alpha = \frac{e(H)}{e(G)} \quad \hat{h}_H = \min_{U \in V(H)} \frac{e(U, \bar{U}) \text{ vol}(H)}{\text{vol}(U) \text{ vol}(\bar{U})}$$

Then

$$\alpha > \hat{h}_H \Rightarrow \forall \lambda \in \text{OPT}(G) \quad H \text{ split}$$

$$\alpha < \hat{h}_H \Rightarrow \forall \lambda \in \text{OPT}(G) \quad H \text{ not split}$$

$$\alpha = \hat{h}_H \Rightarrow \exists \lambda, \lambda' \in \text{OPT}(G)$$

in $\lambda \quad H \text{ split}$

in $\lambda' \quad H \text{ not split}$

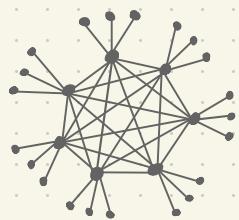
Open

Upper Bound for Modularity
in terms of conductance h_G

Cor $\forall G: q^*(G) \leq 1 - \min\{\hat{h}_G, 1\} \leq 1 - h_G$

Tight for \hat{h}_G : $\forall \hat{\delta} \ 0 < \hat{\delta} \leq 1 \ \forall \varepsilon > 0 \ \exists G$ s.t.

$$|\hat{h}_G - \hat{\delta}| < \varepsilon \quad q^*(G) > 1 - \hat{h}_G - \varepsilon$$



Construction

G k -clique, $k \geq 1$ leaves @ each v

$$\hat{h}_G = \frac{k}{2k+k-1} \quad q^*(G) \geq 1 - \hat{h}_G - o_k(1)$$

Open

What is the optimal f s.t.

$\forall G: q^*(G) \leq 1 - f(h_G) ?$

By



construction $x \leq f(x) \leq 2x$

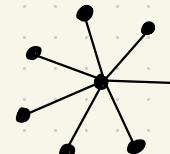
$$q^*(G) \geq 1 - 2h_G - o_k(1)$$

Recall

$$h_G = \min_{U \subseteq V(G)} \frac{e(U, \bar{U})}{\min\{\text{vol}(U), \text{vol}(\bar{U})\}}$$

$$\hat{h}_G = \min_{U \subseteq V(G)} \frac{e(U, \bar{U}) \cdot \text{vol}(G)}{\text{vol}(U) \cdot \text{vol}(\bar{U})}$$

$$\delta = 1$$



$$q^* = 0 \\ \Rightarrow f(1) = 1$$

$$\delta \leq \varepsilon$$



$$q^* \geq 1 - \varepsilon \\ \Rightarrow f(0) = 0$$