Modularity and Graph Expansion

Fiona Skerman

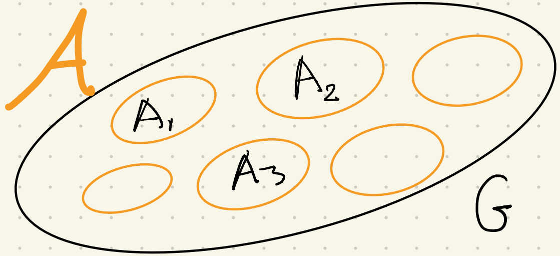
with Baptiste Louf + Colin McDiarmid

arXiv: 2312.07521

Modularity

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

NEWMAN + GIRVAN 2004.



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

score of partition $\mathcal{A}$, $q_{\mathcal{A}}(G) =$

modularity of G $\quad q^*(G) = \max_{\mathcal{A}} q_{\mathcal{A}}(G)$

"high vals taken to indicate more community structure"

Community Detection

input graph $G = (V, E)$

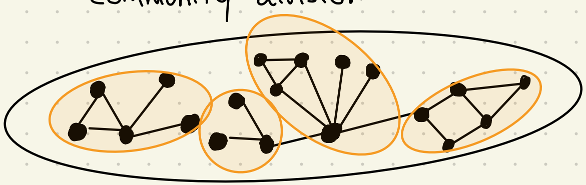
vertices nodes

edges (weighted)

output vertex partition $\mathcal{A}$

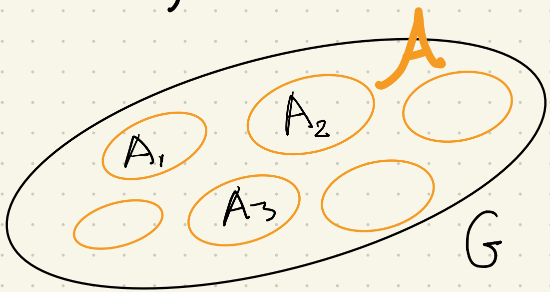
'community division'

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. $\mathcal{A} = \{A_1, \dots, A_k\}$ vertex partition

score of partition $\mathcal{A}$, $q_{\mathcal{A}}(G) =$

modularity of G $\quad q^*(G) = \max_{\mathcal{A}} q_{\mathcal{A}}(G)$

"high vals taken to indicate more community structure"

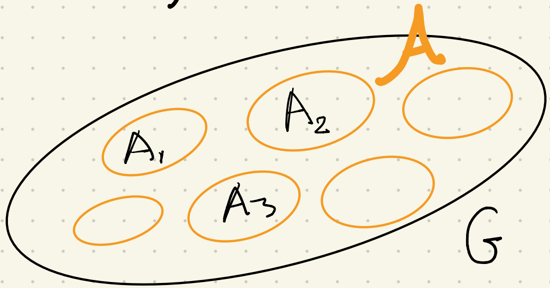
$$q_{\mathcal{A}}(G) = \sum_{A \in \mathcal{A}} \underbrace{\frac{e(A)}{m}}_{\text{"edge contrib"}} - \underbrace{\frac{\mathrm{vol}(A)^2}{(2m)^2}}_{\text{"degree tax"}} = \frac{1}{2m} \sum_{A \in \mathcal{A}} \sum_{u,v \in A} \mathbb{1}[u \sim v] - \frac{d_u d_v}{2m}$$

d_u = # edges incident to u

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

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

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



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

score of partition $\mathcal{A}$, $q_{\mathcal{A}}(G) =$

modularity of G $\quad q^*(G) = \max_{\mathcal{A}} q_{\mathcal{A}}(G)$

"high vals taken to indicate more community structure"

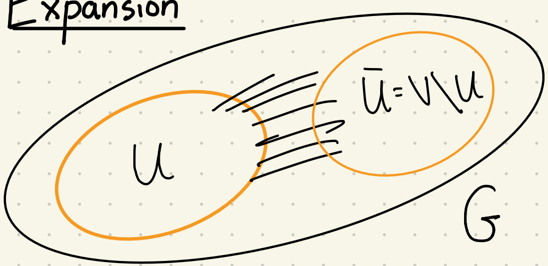
$$q_{\mathcal{A}}(G) = \underbrace{\sum_{A \in \mathcal{A}} \frac{e(A)}{m}}_{\text{"edge contrib"}} - \underbrace{\frac{\mathrm{vol}(A)^2}{(2m)^2}}_{\text{"degree tax"}} = \frac{1}{2m} \sum_{A \in \mathcal{A}} \sum_{u,v \in A} \mathbb{1}[u \sim v] - \frac{d_u d_v}{2m}$$

d_u = # edges incident to u

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

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

Expansion



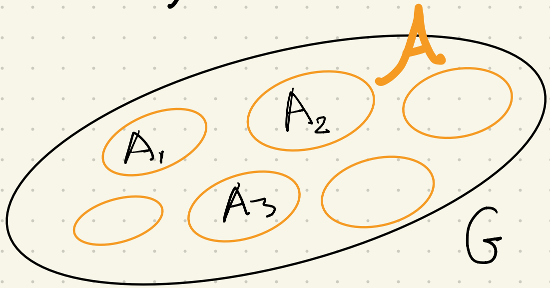
conductance

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

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

$$h_G \leq \hat{h}_G \leq 2 h_G$$

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



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

modularity of G $\qquad q^*(G) = \max_{\mathcal{A}} q_{\mathcal{A}}(G)$

"high vals taken to indicate more community structure"

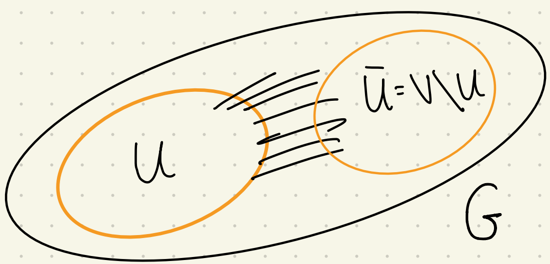
$d_u = \#$edges incident to u

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

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

$$q_{\mathcal{A}}(G) = \underbrace{\sum_{A \in \mathcal{A}} \frac{e(A)}{m}}_{\text{"edge contrib"}} - \underbrace{\frac{vol(A)^2}{(2m)^2}}_{\text{"degree tax"}} = \frac{1}{2m} \sum_{A \in \mathcal{A}} \sum_{u,v \in A} \mathbb{1}[u \sim v] - \frac{d_u d_v}{2m}$$

Expansion



conductance

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

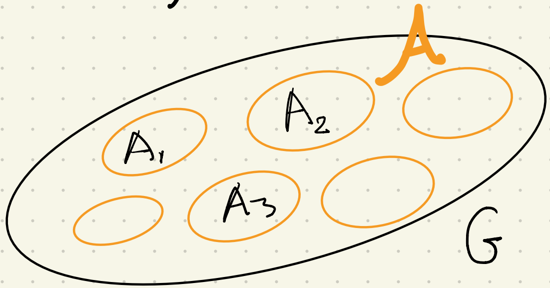
$$\hat{h}_G = \min_{U \subseteq V(G)} \frac{e(U, \bar{U})\, vol(G)}{vol(U)\, vol(\bar{U})}$$

Robustness

$$|q^*(G+1) - q^*(G)| < \frac{2}{e(G)}$$

but $\quad h_{G+1} = \hat{h}_{G+1} = 0 \quad$ disconnected!

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



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

modularity of G $\quad q^*(G) = \max_{\mathcal{A}} q_{\mathcal{A}}(G)$

"high vals taken to indicate more community structure"

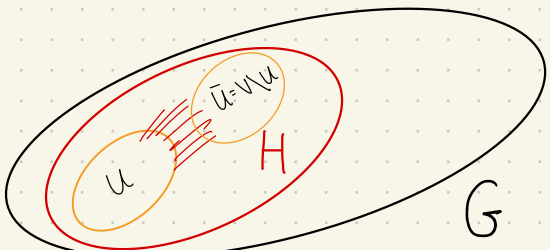
d_u = # edges incident to u

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

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

$$q_{\mathcal{A}}(G) = \underbrace{\sum_{A \in \mathcal{A}} \frac{e(A)}{m}}_{\text{"edge contrib"}} - \underbrace{\frac{\mathrm{vol}(A)^2}{(2m)^2}}_{\text{"degree tax"}} = \frac{1}{2m} \sum_{A \in \mathcal{A}} \sum_{u,v \in A} \mathbb{1}[u \sim v] - \frac{d_u d_v}{2m}$$

Expansion of Subgraphs



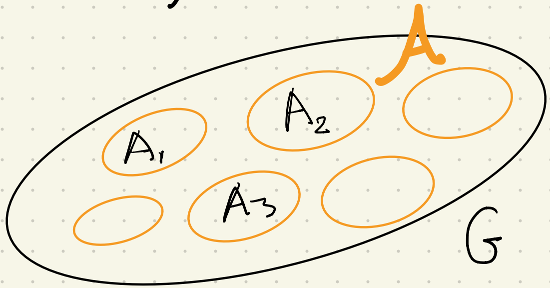
$$\alpha = \frac{e(H)}{e(G)}$$

conductance

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

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

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



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

modularity of G $\quad q^*(G) = \max\limits_{\mathcal{A}} q_{\mathcal{A}}(G)$

"high vals taken to indicate more community structure"

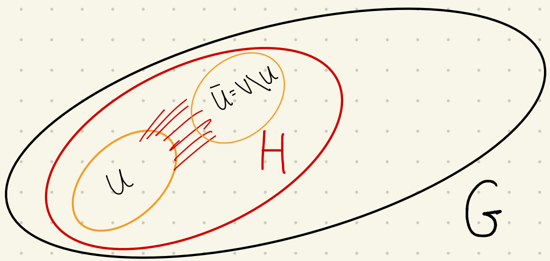
d_u = # edges incident to u

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

$\mathrm{vol}(A) = \sum\limits_{u \in A} d_u$

$$q_{\mathcal{A}}(G) = \underbrace{\sum_{A \in \mathcal{A}} \frac{e(A)}{m}}_{\text{"edge contrib"}} - \underbrace{\frac{\mathrm{vol}(A)^2}{(2m)^2}}_{\text{"degree tax"}} = \frac{1}{2m} \sum_{A \in \mathcal{A}} \sum_{u,v \in A} \mathbb{1}[u \sim v] - \frac{d_u d_v}{2m}$$

Expansion of Subgraphs



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

conductance

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

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

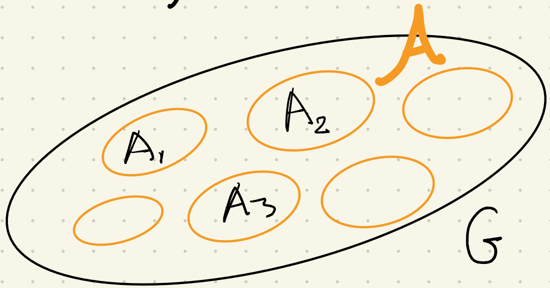
Thm (informal)

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

$\Leftrightarrow$

G has no large expander subgraphs

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



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

modularity of G $\quad q^*(G) = \max_{\mathcal{A}} q_{\mathcal{A}}(G)$

"high vals taken to indicate more community structure"

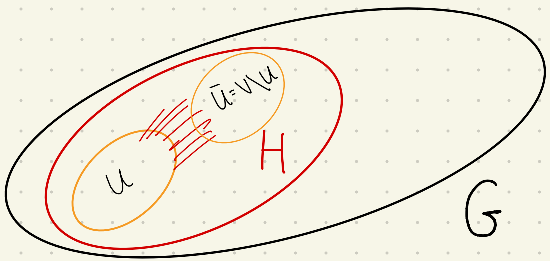
d_u = # edges incident to u

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

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

$$q_{\mathcal{A}}(G) = \underbrace{\sum_{A \in \mathcal{A}} \frac{e(A)}{m}}_{\text{"edge contrib"}} - \underbrace{\frac{\mathrm{vol}(A)^2}{(2m)^2}}_{\text{"degree tax"}} = \frac{1}{2m} \sum_{A \in \mathcal{A}} \sum_{u,v \in A} \mathbb{1}[u \sim v] - \frac{d_u d_v}{2m}$$

Expansion of Subgraphs



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

conductance

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

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

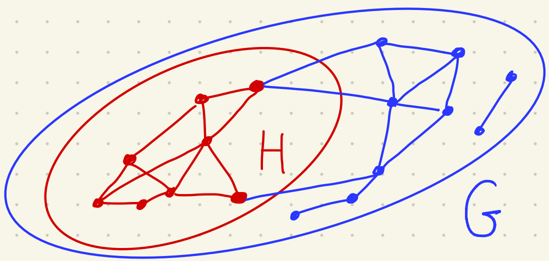
Thm (informal)

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

<=>

G has no large expander subgraphs

Modularity + Expansion of Subgraphs



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

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

"edge contrib." "degree tax"

$$q^*(G) = \max_{\mathcal{A}} q_{\mathcal{A}}(G).$$

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

d_u = # edges incident to u

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

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

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

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

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

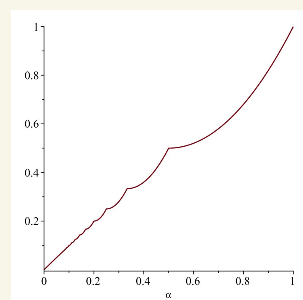
(b) $\exists \delta > 0$: $q^*(G) \leq 1 - f(\alpha) - \varepsilon$

$\Rightarrow$ G has induced subgraph H,

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

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

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



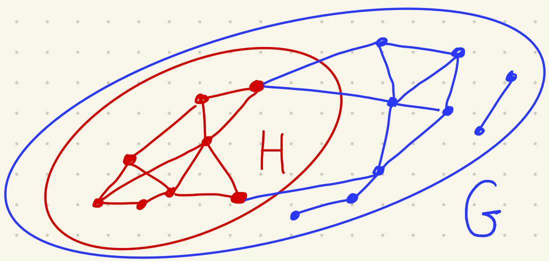
Thm (informal)

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

$\Leftrightarrow$

G has no large expander subgraphs

Modularity + Expansion of Subgraphs



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

$$q_{\mathcal{A}}(G) = \sum_{A \in \mathcal{A}} \frac{e(A)}{m} - \frac{\mathrm{vol}(A)^2}{(2m)^2}$$

"edge contrib" "degree tax"

$$q^*(G) = \max_{\mathcal{A}} q_{\mathcal{A}}(G).$$

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

d_u = # edges incident to u

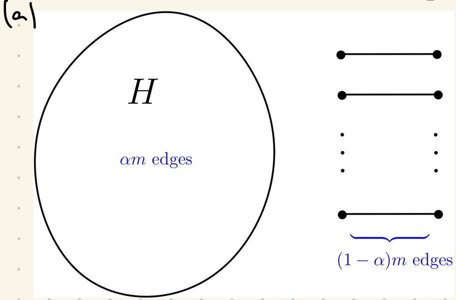
$e(A)$ = # edges in set A

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

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

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

(a)



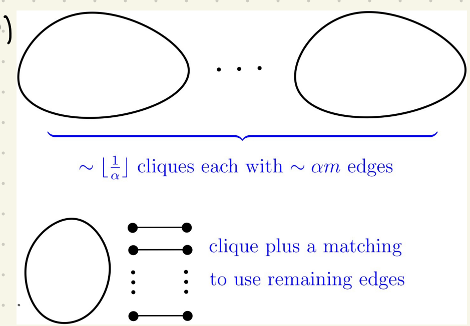
$q^*(G_H) > 1 - \alpha^2 - \varepsilon$

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

$\Rightarrow$ G has induced subgraph H,

$\dfrac{e(H)}{e(G)} \geq \alpha$, $h_H \geq \delta$.

(b)



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

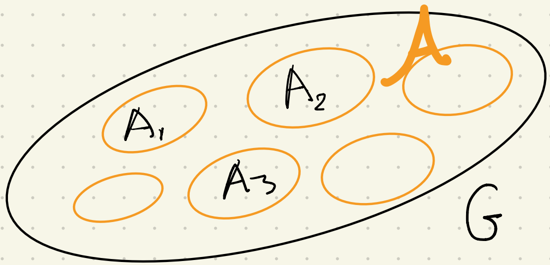
any H, with $\dfrac{e(H)}{e(G)} \geq \alpha$ is disconnected

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

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

Modularity

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



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

score of partition $\mathcal{A}$, $q_{\mathcal{A}}(G) =$

modularity of G $\quad q^*(G) = \max_{\mathcal{A}} q_{\mathcal{A}}(G)$

"high vals taken to indicate more community structure"

$$q_{\mathcal{A}}(G) = \sum_{A \in \mathcal{A}} \underbrace{\frac{e(A)}{m}}_{\text{"edge contrib"}} - \underbrace{\frac{vol(A)^2}{(2m)^2}}_{\text{"degree tax"}} = \frac{1}{2m} \sum_{A \in \mathcal{A}} \sum_{u,v \in A} \mathbb{1}[u \sim v] - \frac{d_u d_v}{2m}$$

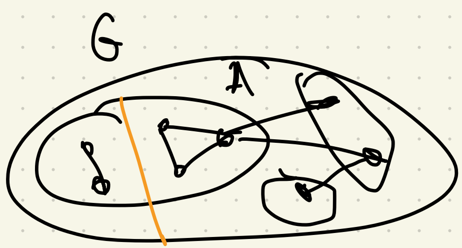
d_u = # edges incident to u

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

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

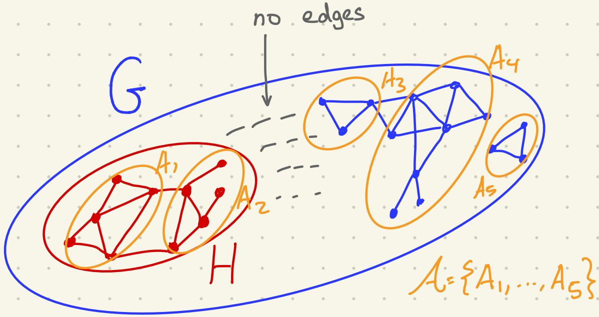
Obs

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



Modularity: Resolution Limit

OPT(G) : set of partitions achieving maximal modularity

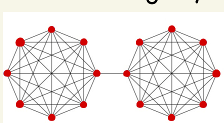


FORTUNATO + BARTHÉLEMY 2008.

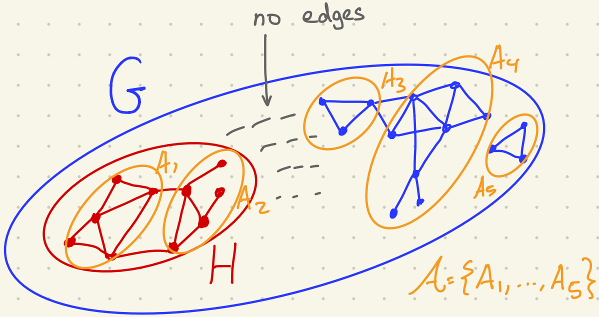
Res Limit

H is a connected component in graph G.

- $\forall H$: $e(H) < \sqrt{2\, e(G)}$ ⇒ H not split
- H 'dumbell graph' $e(H) > \sqrt{2\, e(G)}$ ⇒ H split



Modularity: Resolution Limit

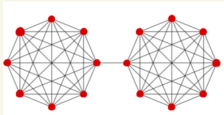


FORTUNATO + BARTHÉLEMY 2008.

Res Limit

H is a connected component in graph G.

- $\forall H$: $e(H) < \sqrt{2\, e(G)} \Rightarrow H$ not split
- H 'dumbell graph' $e(H) > \sqrt{2\, e(G)} \Rightarrow H$ split



$OPT(G)$: set of partitions achieving maximal modularity

Thm H is a connected component in graph G. Let

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

Then

$\alpha > \hat{h}_H \Rightarrow \forall \mathcal{A} \in OPT(G)$ H split

$\alpha < \hat{h}_H \Rightarrow \forall \mathcal{A} \in OPT(G)$ H not split

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

in $\mathcal{A}$ H split

in $\mathcal{A}'$ H not split

Open

Upper Bound for Modularity in terms of conductance h_G

Recall

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

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

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

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

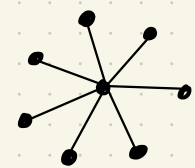
$|\hat{h}_G - \hat{\delta}| < \varepsilon$ $\quad q^*(G) > 1 - \hat{h}_G - \varepsilon$

Construction

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

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

$\delta = 1$



$q^* = 0$

$\Rightarrow f(1) = 1$

$\delta \leq \varepsilon$

$q^* \geq 1 - \varepsilon$

$\Rightarrow f(0) = 0$

Open

What is the optimal f s.t
$\forall G$: $\quad q^*(G) \leq 1 - f(h_G)$?

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

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