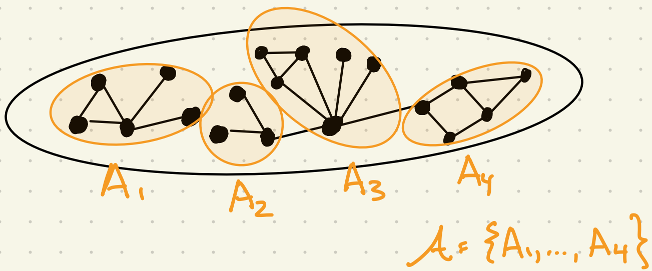
Modularity: a redemption?, resolution limit and graph expansion

Fiona Skerman, Vilhelm Agdur, Nina Kamčev, Baptiste Louf, Colin McDiarmid




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

G



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

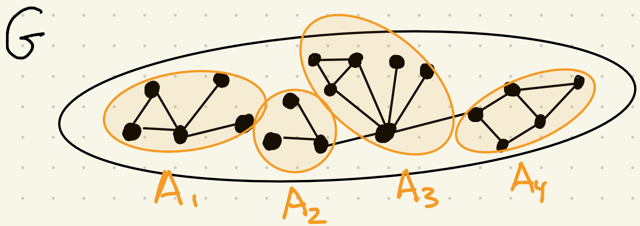
score of partition $\mathcal{A}$, $q_{\mathcal{A}}(G) =$

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

"high vals taken to indicate more community structure"

Modularity

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



$\mathcal{A} = \{A_1, \dots, A_4\}$

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 $G = (V, E)$
- V: vertices / nodes
- E: edges (weighted)

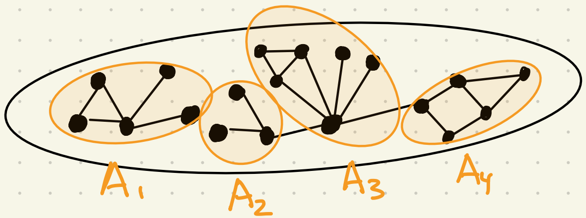
output vertex partition $\mathcal{A}$
'community division'

- modularity score NP-hard to opt.
- Louvain & Leiden ~ 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'

G



$\mathcal{A} = \{A_1, \dots, A_4\}$

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

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

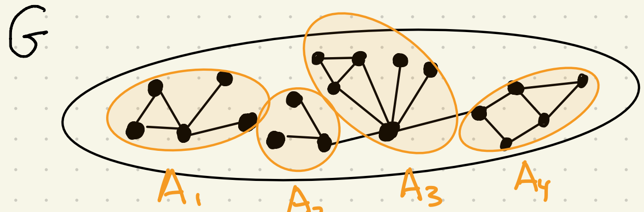
$\searrow \frac{|A|^2}{n^2}$ for regular graphs

$d_u = \#$edges incident to u

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

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

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



$\mathcal{A} = \{A_1, \dots, A_4\}$

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"

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

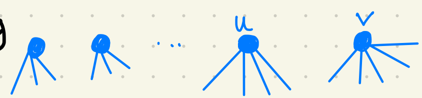
"edge contrib." "degree tax"

d_u = # edges incident to u

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

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

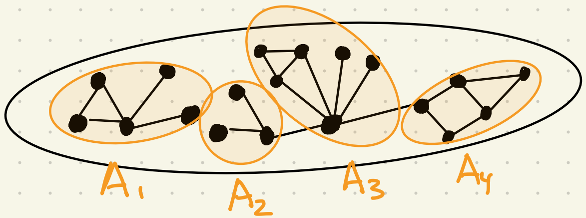
$R \sim$ config model



$$q_{\mathcal{A}}(G) \approx \frac{1}{m}\left(e^{int}_G(\mathcal{A}) - \mathbb{E}\left[\, e^{int}_R(\mathcal{A})\right]\right)$$

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

G



$\mathcal{A}=\{A_1,\dots,A_4\}$

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"

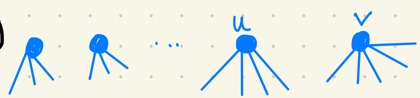
$$q_{\mathcal{A}}(G) = \sum_{A\in\mathcal{A}} \underbrace{\frac{e(A)}{m}}_{\text{"edge contrib"}} - \underbrace{\frac{vol(A)^2}{(2m)^2}}_{\text{"degree tax"}} = \frac{1}{2m}\sum_{A\in\mathcal{A}}\sum_{u,v\in A} \mathbb{1}[u\sim v] - \frac{d_u d_v}{2m}$$

d_u = # edges incident to u

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

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

config model

$R \sim$ 

$$q_{\mathcal{A}}(G)\approx \frac{1}{m}\left(e^{int}_G(\mathcal{A}) - \mathbb{E}[e^{int}_R(\mathcal{A})]\right)$$

- $u\neq v \quad \mathbb{E}[\#\text{edges } u\sim v \text{ in } R] = \frac{d_u d_v}{2m-1}$
- $\mathbb{E}[\#\text{edges within parts of } \mathcal{A} \text{ in } R] = \sum_{A\in\mathcal{A}} \frac{vol(A)(vol(A)-1)}{2m(2m-1)}$

Example : Linguistics

$V \approx 3000$ climate change blogs

$E \sim$ based on links between blogs

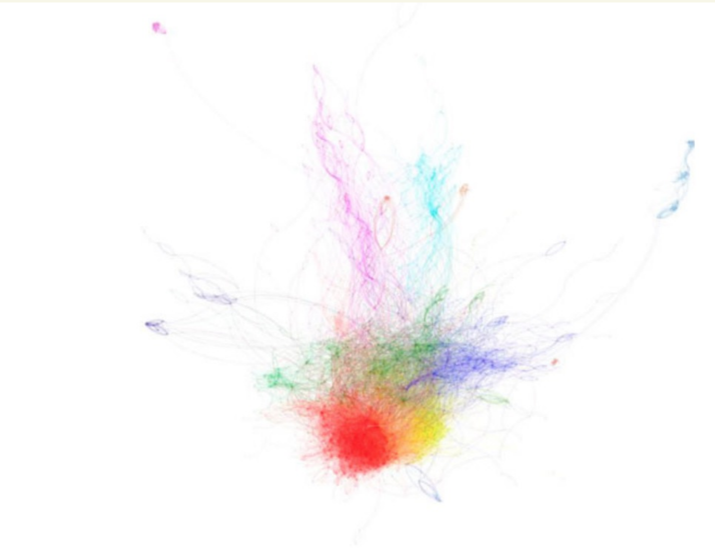


Figure 1. The network of climate change blogs, colored by community.

Elgesam D, Steskal L. + Diakopoulos

"Structure and content of the discourse on climate change in the blogosphere"

Environmental Communication '2015.

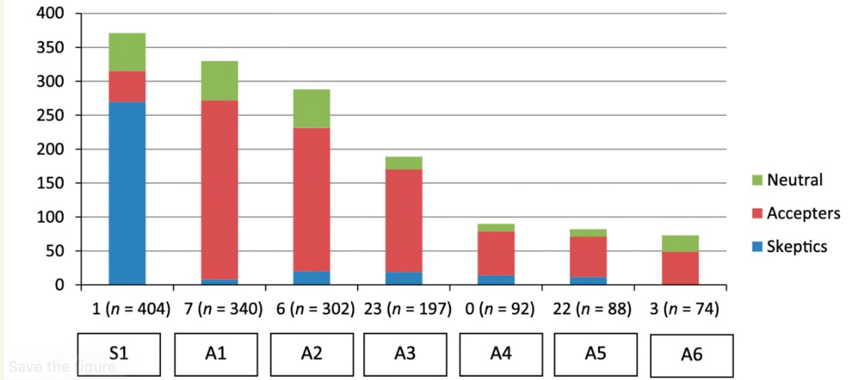


Figure 3. The distribution of skeptical, accepting, and neutral blogs in the seven largest among the central groups of blogs concerned with climate change.

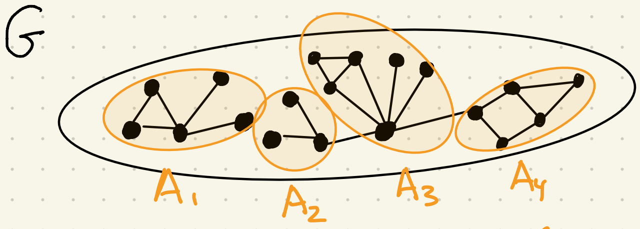
Table 5. The top 15 collocates around "climate" in communities 1 (skeptic), 23 (accepter), and 7 (accepter) computed with the point-wise mutual information metric.

Top collocates of "CLIMATE" in the skeptical community S1	Top collocates of "CLIMATE" in the accepter community A3	Top collocates of "CLIMATE" in the accepter community A1
1 CLIMATE	1 DENIERS	1 POPPIN
2 SKEPTICS	2 SKEPTICS	2 DENIERS
3 ALARMISM	3 CLIMAT	3 SKEPTICS
4 DENIERS	4 DECADAL	4 OBAMA
5 IPCC	5 CONTRARIANS	5 WWW
6 DECADAL	6 OBAMA	6 EU'S
7 ALARMISTS	7 NOAA'S	7 CLIMATE
8 CLIMAT	8 AGW	8 YVO
9 CHANGE	9 WWW	9 NOAA'S
10 INTERGOVERNMENTAL	10 DENIER	10 WILDFIRES
11 OBAMA	11 CLIMATE	11 CHANGE'S
12 ANTHROPOGENIC	12 VAPOR	12 IPCC
13 AGW	13 ANTHROPOGENIC	13 ALARMISM
14 IPCC'S	14 ALARMISM	14 PACHAURI
15 WARMING	15 CONTRARIAN	15 DENIER

Reference corpus: The British National Corpus, approximately 100 million words.

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"

$\mathcal{A} = \{A_1, \dots, A_4\}$

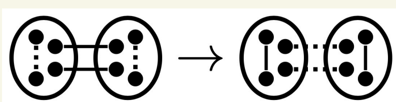
$$q_{\mathcal{A}}(G) = \sum_{A \in \mathcal{A}} \underbrace{\frac{e(A)}{m}}_{\text{"edge contrib"}} - \underbrace{\frac{\mathrm{vol}(A)^2}{(2m)^2}}_{\text{"degree tax"}} = \frac{1}{2m} \sum_{A \in \mathcal{A}} \sum_{u,v \in A} \mathbb{1}[u \sim v] - \frac{d_u d_v}{2m}$$

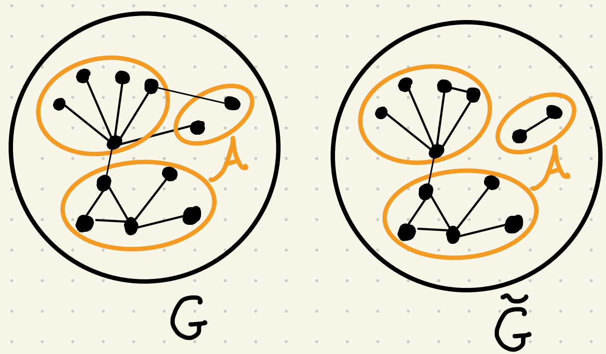
d_u = # edges incident to u

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

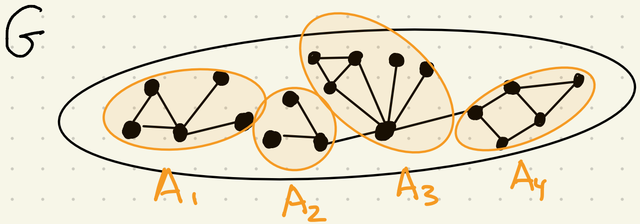
$\mathrm{vol}(A) = \sum_{u \in A} d_u$

Fix $\mathcal{A}$, which $G \to \tilde{G}$ ensures $q_{\mathcal{A}}(\tilde{G}) > q_{\mathcal{A}}(G)$?





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



$\mathcal{A} = \{A_1, \ldots, A_4\}$

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"

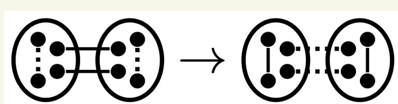
$$q_{\mathcal{A}}(G) = \underbrace{\sum_{A \in \mathcal{A}} \frac{e(A)}{m}}_{\text{"edge contrib"}} - \underbrace{\frac{vol(A)^2}{(2m)^2}}_{\text{"degree tax"}} = \frac{1}{2m} \sum_{A \in \mathcal{A}} \sum_{u,v \in A} \mathbb{1}[u \sim v] - \frac{d_u d_v}{2m}$$

d_u = # edges incident to u

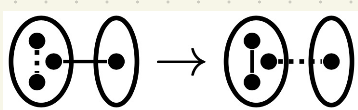
$e(A)$ = # edges in set A

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

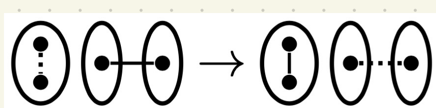
Fix $\mathcal{A}$, which $G \to \tilde{G}$ ensures $q_{\mathcal{A}}(\tilde{G}) > q_{\mathcal{A}}(G)$?



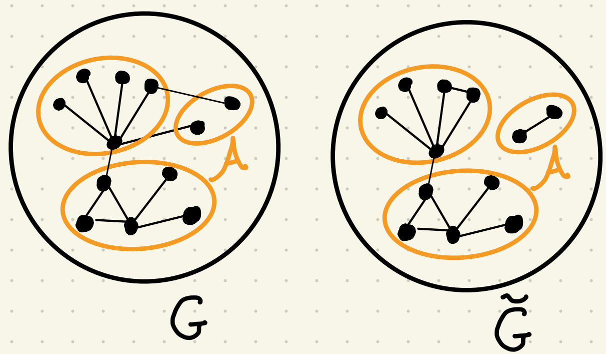
✓

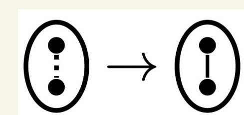
* 

✓

* 

✗





✗

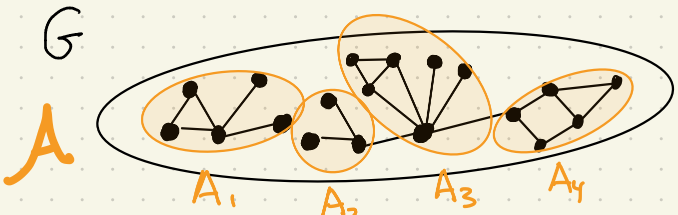
✗

(✓ if $|\mathcal{A}| = 2$, $q_{\mathcal{A}}(G) \geq 0$)

*TRAJANOVSKI, WANG, VAN MIEGHEM '12.

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"

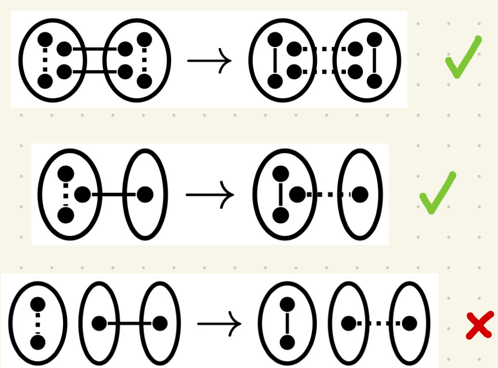
$$q_A(G) = \underbrace{\sum_{A \in \mathcal{A}} \frac{e(A)}{m}}_{\text{"edge contrib"}} - \underbrace{\frac{\mathrm{vol}(A)^2}{(2m)^2}}_{\text{"degree tax"}} = \frac{1}{2m} \sum_{A \in \mathcal{A}} \sum_{u,v \in A} \mathbb{1}[u \sim v] - \frac{d_u d_v}{2m}$$

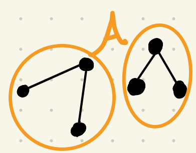
$d_u = \#$edges incident to u

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

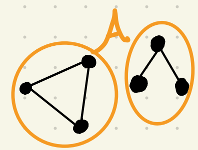
$\mathrm{vol}(A) = \sum_{u \in A} d_u$

Fix A, which $G \to \tilde{G}$ ensures $q_A(\tilde{G}) > q_A(G)$?

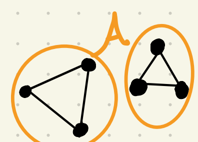




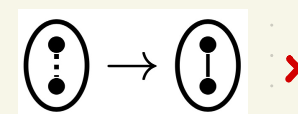
$q_A = \frac{1}{2}$

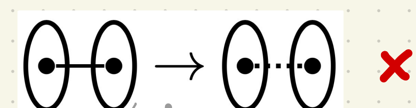


$q_A = \frac{12}{25}$



$q_A = \frac{1}{2}$

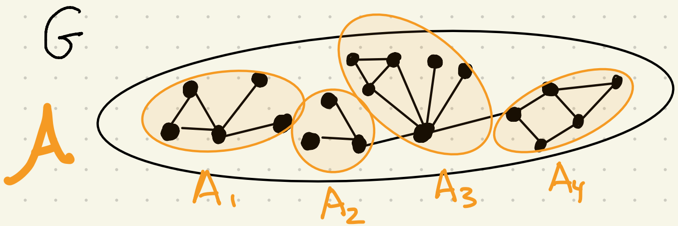




(✓ if $|A| = 2$, $q_A(G) \geq 0$)

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

"high vals taken to indicate more community structure"

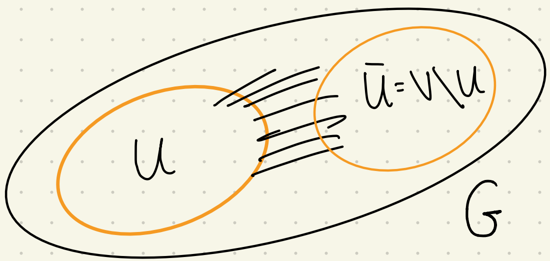
d_u = # edges incident to u

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

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

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

Expansion



conductance

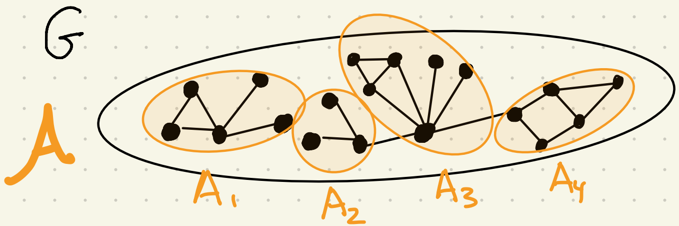
$$h_G = \min_{U \subseteq V(G)} \frac{e(U, \bar{U})}{\min\{\mathrm{vol}(U), \mathrm{vol}(\bar{U})\}}$$

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

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

Modularity

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



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

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

"high vals taken to indicate more community structure"

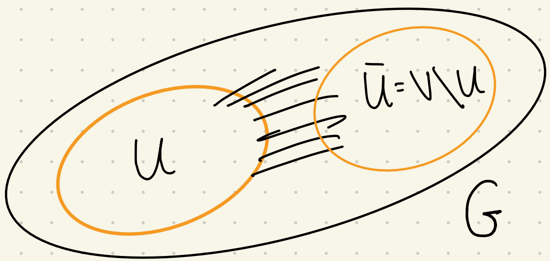
$$q_{\mathcal{A}}(G) = \sum_{A\in\mathcal{A}} \underbrace{\frac{e(A)}{m}}_{\text{"edge contrib"}} - \underbrace{\frac{vol(A)^2}{(2m)^2}}_{\text{"degree tax"}} = \frac{1}{2m}\sum_{A\in\mathcal{A}}\sum_{u,v\in A} \mathbb{1}[u\sim v] - \frac{d_u d_v}{2m}$$

$d_u = \#$ edges incident to u

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

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

Expansion



conductance

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

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

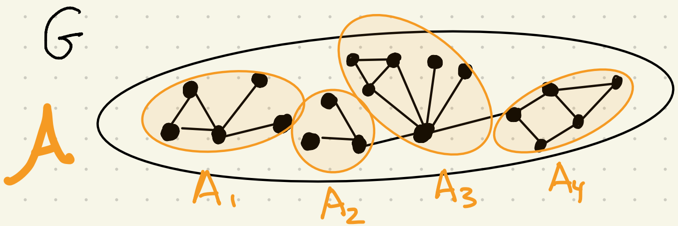
Robustness

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

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

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

"high vals taken to indicate more community structure"

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

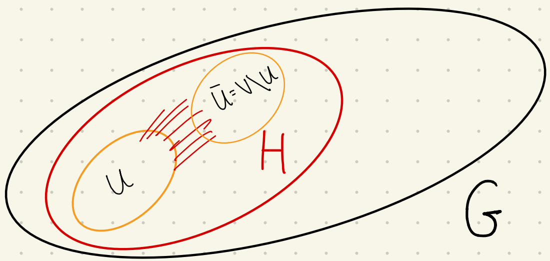
"edge contrib." "degree tax"

d_u = # edges incident to u

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

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

Expansion of Subgraphs



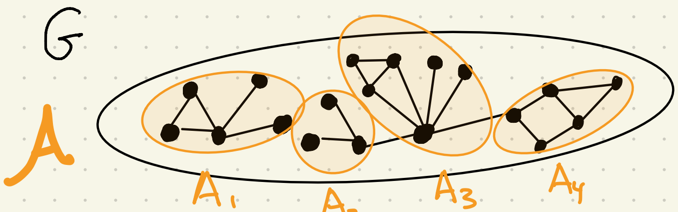
$$\alpha = \frac{e(H)}{e(G)}$$

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

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

"high vals taken to indicate more community structure"

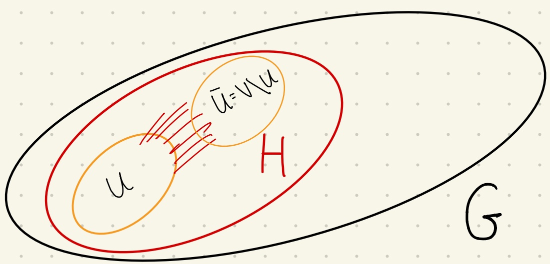
d_u = # edges incident to u

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

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

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

Expansion of Subgraphs



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

conductance

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

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

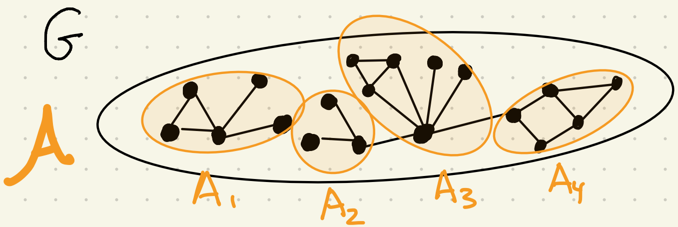
Thm (informal)

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"

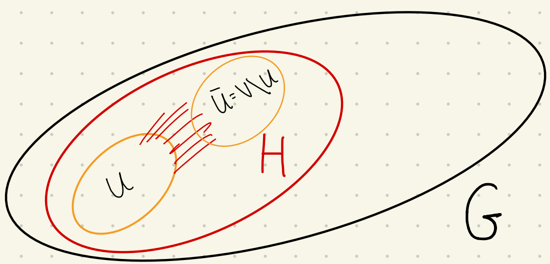
$$q_{\mathcal{A}}(G) = \sum_{A \in \mathcal{A}} \underbrace{\frac{e(A)}{m}}_{\text{"edge contrib."}} - \underbrace{\frac{vol(A)^2}{(2m)^2}}_{\text{"degree tax"}} = \frac{1}{2m} \sum_{A \in \mathcal{A}} \sum_{u,v \in A} \mathbb{1}[u \sim v] - \frac{d_u d_v}{2m}$$

d_u = # edges incident to u

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

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

Expansion of Subgraphs



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

conductance

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

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

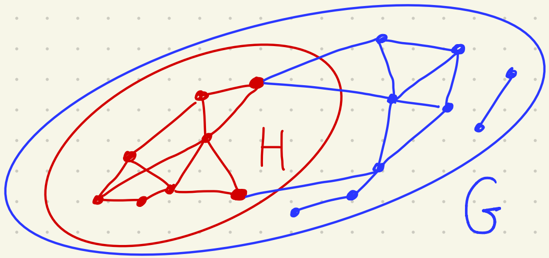
Thm (informal)

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

<=>

G has no large expander subgraphs

Modularity + Expansion of Subgraphs



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

Thm (informal)

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

<=>

G has no large expander subgraphs

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

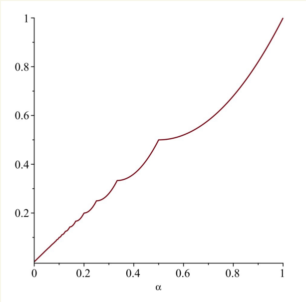
(a) G has subgraph H, $\frac{e(H)}{e(G)} \geq \alpha$, $h_H \geq \alpha$

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

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

$\Rightarrow$ G has induced subgraph H,

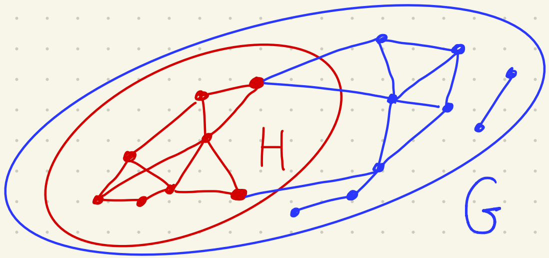
$\frac{e(H)}{e(G)} \geq \alpha$, $h_H \geq \delta$.



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

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

Modularity + Expansion of Subgraphs



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

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

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

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

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

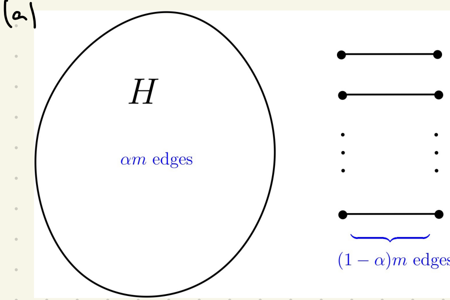
$\Rightarrow$ G has induced subgraph H,

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

$f(\alpha) := \max\{\sum_i x_i^2 : 0 \leq x_i \leq \alpha, \sum_1 x_i = 1\}$

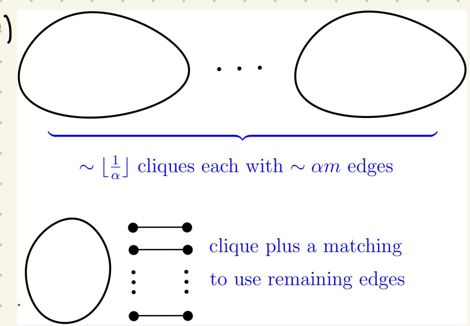
N.B. $f(\alpha) \approx \alpha$ for small α

(a)



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

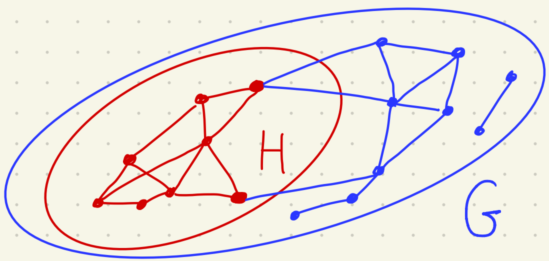
(b)



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

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

Modularity + Expansion of Subgraphs



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

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

$\hat{h}_H = \frac{k}{2\ell + k - 1}$

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

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

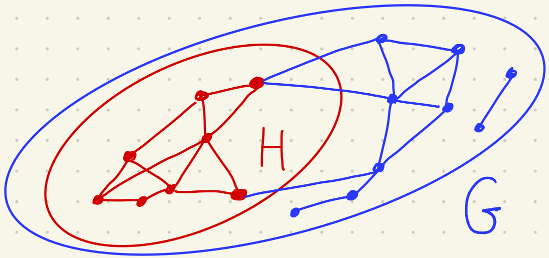
$$\Rightarrow q^*(G) \leq 1 - \alpha \min\{\alpha, \hat{h}_H\}$$

$\forall \varepsilon > 0$ $\forall \hat{\delta}, a$ $\quad 0 \leq a, \hat{\delta} \leq 1$

$\exists G, H$ st. $\quad |\hat{h}_H - \hat{\delta}|, |a - \alpha| < \varepsilon$

$$q^*(G) \geq 1 - \alpha \min\{\alpha, \hat{h}_H\} - \varepsilon$$

Modularity + Expansion of Subgraphs



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

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

$$\hat{h}_H = \frac{k}{2\ell + k - 1}$$

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

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

$$\Rightarrow \quad q^*(G) \leq 1 - \alpha \min\{\alpha, \hat{h}_H\}$$

$\forall \varepsilon > 0$ $\forall \hat{\delta}, a$ $\quad 0 \leq a, \hat{\delta} \leq 1$

$\exists G, H$ st. $\quad |\hat{h}_H - \hat{\delta}|, |a - \alpha| < \varepsilon$

$$q^*(G) \geq 1 - \alpha \min\{\alpha, \hat{h}_H\} - \varepsilon$$

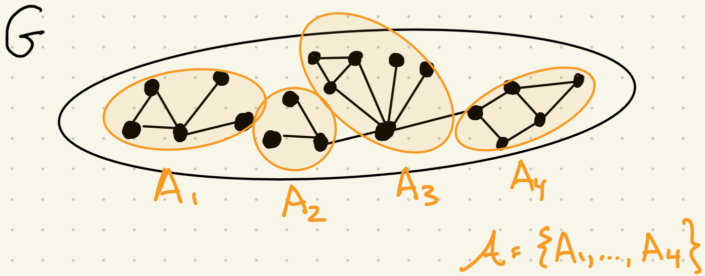
(b) $\forall \alpha, \delta > 0$: any H an induced subgraph of G with $\frac{e(H)}{e(G)} \geq \alpha$ has $h_H \leq \delta$.

$$\Rightarrow \quad q^*(G) \geq 1 - f(\min\{1, \alpha + \tfrac{3}{2}\delta\}) - \tfrac{3}{2}\delta \lceil \log_2(\alpha^{-1}) \rceil$$

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

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"

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

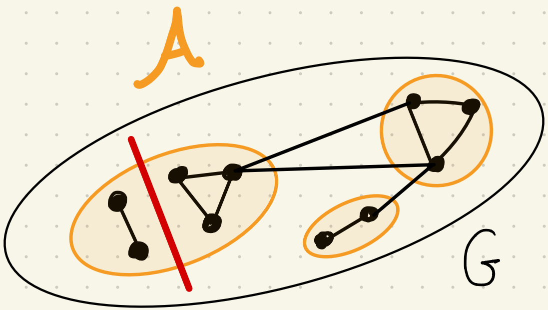
$d_u = \#$edges incident to u

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

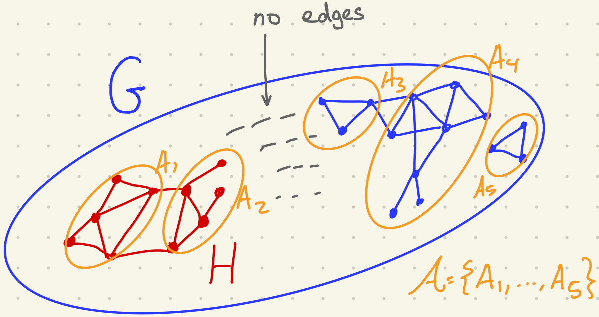
$vol(A) = \sum_{u \in A} d_u$

Obs $\quad \mathcal{A}$ optimal partition of G i.e $q_{\mathcal{A}}(G) = q^*(G)$

$\Rightarrow \forall A \in \mathcal{A} \quad G[A]$ conn. $\quad$ (+ isolated vert)



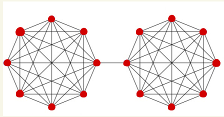
Resolution Limit



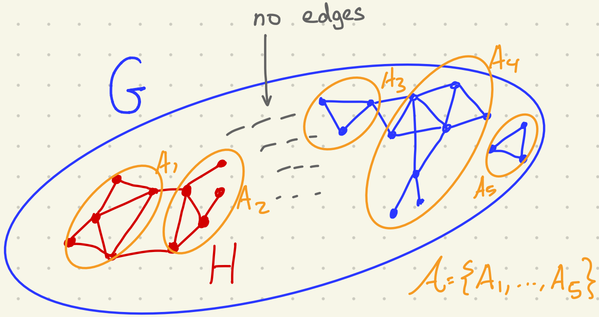
FORTUNATO + BARTHÉLEMY 2007.

H is a connected component in graph G.

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



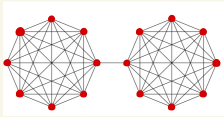
Resolution Limit



FORTUNATO + BARTHÉLEMY 2007.

H is a connected component in graph G.

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



Let

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

conductance :

$$h_H = \min_{U \subseteq V(H)} \frac{e(U, \bar{U})}{\min\{vol(U), vol(\bar{U})\}}. \qquad h_H \leq \hat{h}_H \leq 2 h_H$$

Thm

H is a connected component in graph G

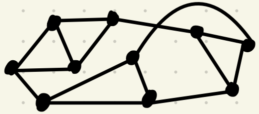
Then $\forall H$: $\quad e(H) < \hat{h}_H\, e(G) \Rightarrow H$ not split

$\quad e(H) > \hat{h}_H\, e(G) \Rightarrow H$ split

Modularity : A Redemption?

- PROBLEM: RANDOM GRAPHS HAVE HIGH MODULARITY

RANDOM d-REG



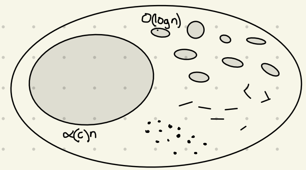
$\exists c$ s.t. whp

$$c/\sqrt{d} < q^*(G_{n,d}) < 2/\sqrt{d}$$

[LICHEV + MITSCHE] whp

$$0.667 < q^*(G_{n,3}) < 0.789998$$

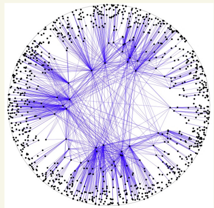
ERDÖS-RENYI



$\frac{1}{n} < p < 0.99$. whp

$$q^*(G_{n,p}) = \Theta\left(\frac{1}{\sqrt{np}}\right) = \Theta\left(\frac{1}{\sqrt{d}}\right)$$

RANDOM HYPERBOLIC



$$q^*(G_n) = 1 - o(1) \quad \text{whp}$$

PHYSICAL REVIEW E **70**, 025101(R) (2004)

Modularity from fluctuations in random graphs and complex networks

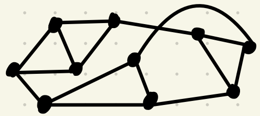
Roger Guimerà, Marta Sales-Pardo, and Luís A. Nunes Amaral
Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, USA
(Received 15 March 2004; published 19 August 2004)

A differentiation between graphs which are truly modular and those which are not can … only be made if we gain an understanding of the intrinsic modularity of random graphs. – Reichardt and Bornholdt

Modularity : A Redemption?

- PROBLEM: RANDOM GRAPHS HAVE HIGH MODULARITY

RANDOM d-REG



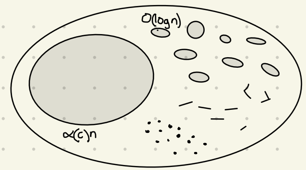
$\exists c$ s.t. whp

$$\frac{c}{\sqrt{d}} < q^*(G_{n,d}) < 2/\sqrt{d}$$

[LICHEV + MITSCHE] whp

$$0.667 < q^*(G_{n,3}) < 0.789998$$

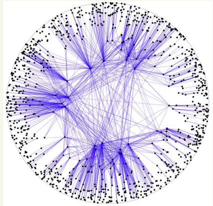
ERDÖS-RENYI



$\frac{1}{n} < p < 0.99$. whp

$$q^*(G_{n,p}) = \Theta\left(\frac{1}{\sqrt{np}}\right) = \Theta\left(\frac{1}{\sqrt{d}}\right)$$

RANDOM HYPERBOLIC



$$q^*(G_n) = 1 - o(1) \quad \text{whp}$$

PHYSICAL REVIEW E **70**, 025101(R) (2004)

Modularity from fluctuations in random graphs and complex networks

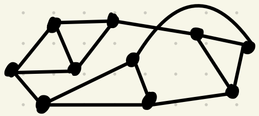
Roger Guimerà, Marta Sales-Pardo, and Luís A. Nunes Amaral
Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, USA
(Received 15 March 2004; published 19 August 2004)

A differentiation between graphs which are truly modular and those which are not can ... only be made if we gain an understanding of the intrinsic modularity of random graphs. – Reichardt and Bornholdt

Modularity : A Redemption?

- PROBLEM: RANDOM GRAPHS HAVE HIGH MODULARITY

RANDOM d-REG



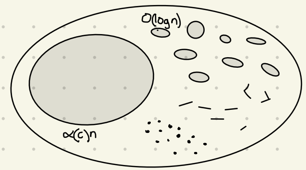
$\exists c$ s.t. whp

$c/\sqrt{d} < q^*(G_{n,d}) < 2/\sqrt{d}$

[LICHEV + MITSCHE] whp

$0.667 < q^*(G_{n,3}) < 0.79$

ERDÖS-RENYI



$\frac{1}{n} < p < 0.99$. whp

$q^*(G_{n,p}) = \Theta\left(\frac{1}{\sqrt{np}}\right) = \Theta\left(\frac{1}{\sqrt{d}}\right)$

CHUNG-LU

$(w_1, \dots, w_n)$ $\bar{w} = \frac{1}{n}\sum_i w_i$ $w_i^2 = o(\bar{w}n)\ \forall i$

$P[ij \text{ edge}] = \frac{w_i w_j}{\bar{w} n}$

$q^*(G_{n,w}) \lesssim 4/\sqrt{\bar{w}}$ $w_{min} \gg \log^2 n$

" $\lesssim c_0/\sqrt{w_{min}}$ $w_{min} = \Omega(1)$

PHYSICAL REVIEW E **70**, 025101(R) (2004)

Modularity from fluctuations in random graphs and complex networks

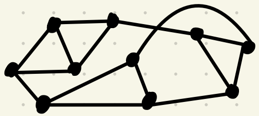
Roger Guimerà, Marta Sales-Pardo, and Luís A. Nunes Amaral
Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, USA
(Received 15 March 2004; published 19 August 2004)

A differentiation between graphs which are truly modular and those which are not can … only be made if we gain an understanding of the intrinsic modularity of random graphs. – Reichardt and Bornholdt

Modularity : A Redemption?

- PROBLEM: RANDOM GRAPHS HAVE HIGH MODULARITY

RANDOM d-REG



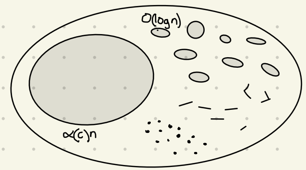
$\exists c$ s.t. whp

$$c/\sqrt{d} < q^*(G_{n,d}) < 2/\sqrt{d}$$

[LICHEV + MITSCHE] whp

$$0.667 < q^*(G_{n,3}) < 0.79$$

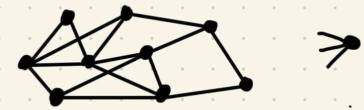
ERDÖS-RENYI



$\frac{1}{n} < p < 0.99.$ whp

$$q^*(G_{n,p}) = \Theta\left(\frac{1}{\sqrt{np}}\right) = \Theta\left(\frac{1}{\sqrt{d}}\right)$$

PREF. ATTACHMENT $(h \geq 2)$



[RYBARCZYK + SULKOWSKA] (upper bound)

whp $q^*(G_n^h) = \tilde{\Theta}\left(\frac{1}{\sqrt{h}}\right)$

PHYSICAL REVIEW E **70**, 025101(R) (2004)

Modularity from fluctuations in random graphs and complex networks

Roger Guimerà, Marta Sales-Pardo, and Luís A. Nunes Amaral
Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, USA
(Received 15 March 2004; published 19 August 2004)

A differentiation between graphs which are truly modular and those which are not can ... only be made if we gain an understanding of the intrinsic modularity of random graphs. – Reichardt and Bornholdt

Modularity : A Redemption?

- PROBLEM: RANDOM GRAPHS HAVE HIGH MODULARITY

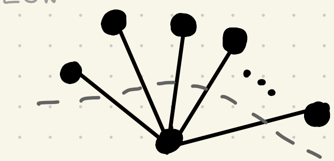
PHYSICAL REVIEW E **70**, 025101(R) (2004)

Modularity from fluctuations in random graphs and complex networks

Roger Guimerà, Marta Sales-Pardo, and Luís A. Nunes Amaral
Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, USA
(Received 15 March 2004; published 19 August 2004)

Low

$\sum_{i \in \text{Low}} d_i = k = \frac{\text{vol}(G)}{2}$



star

$q^* = 0$

k edges

Theorem 1.1. *Let G be an n-vertex graph with average degree $\bar{d} \geq 1$, $L = \{v \in G \mid d_v < C\bar{d}\}$ for some $C > 1$, and assume that $\text{vol}(L) \geq (1+\gamma)m = (1+\gamma)\frac{n\bar{d}}{2}$ for some $\gamma > 0$.*

If $\Delta(G)n^{-1} = o(1)$ and $\bar{d}^{10}n^{-1} = o(1)$, then

$$q^*(G) \geq \frac{0.26\gamma}{\sqrt{C\bar{d}}}(1 + o(1)).$$

Modularity : A Redemption?

- PROBLEM: RANDOM GRAPHS HAVE HIGH MODULARITY

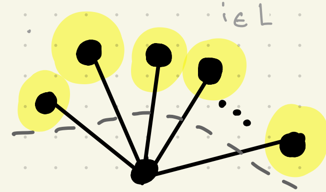
PHYSICAL REVIEW E **70**, 025101(R) (2004)

Modularity from fluctuations in random graphs and complex networks

Roger Guimerà, Marta Sales-Pardo, and Luís A. Nunes Amaral
Department of Chemical and Biological Engineering, Northwestern University, Evanston, Illinois 60208, USA
(Received 15 March 2004; published 19 August 2004)

$\sum_{i \in L} d_i = k = \frac{vol(G)}{2}$

L



star

$q^* = 0$

k edges

Theorem 1.1. *Let G be an n-vertex graph with average degree $\bar{d} \geq 1$, $L = \{v \in G \mid d_v < C\bar{d}\}$ for some $C > 1$, and assume that $\mathrm{vol}(L) \geq (1+\gamma)m = (1+\gamma)\frac{n\bar{d}}{2}$ for some $\gamma > 0$.*

If $\Delta(G)n^{-1} = o(1)$ and $\bar{d}^{10}n^{-1} = o(1)$, then

$$q^*(G) \geq \frac{0.26\gamma}{\sqrt{C\bar{d}}}(1 + o(1)).$$

Proof Idea $\quad q^*(G) = \Omega(d^{-1/2})$, $\quad d = o(n^{1/10})$.

Load balancing $\quad n$ even

$$x_1 \geq x_2 \geq x_3 \geq x_4 \quad \dots \quad \geq x_n$$

$$\underbrace{A \qquad B}\qquad \underbrace{B \qquad A}\qquad \underbrace{A \qquad B}$$

$$\left|\sum_{i\in A} x_i - \sum_{i\in B} x_i\right| \leq \sum_{i \text{ odd}} x_i - \sum_{i \text{ even}} x_i$$

$$= x_{max} - x_2 + x_3 - \dots - x_{min}$$

$$\leq x_{max} - x_{min}$$

$$x_1 \geq x_2 \geq x_3 \geq x_4 \quad \dots \quad \geq x_n$$

A A B B B A A B

$$\longleftarrow \leq \ell \longrightarrow$$

$$\left|\sum_{i\in A} x_i - \sum_{i\in B} x_i\right| \leq \ell(x_{max} - x_{min})$$

Robustness

- <u>Robust</u> to small perturbations in edge set

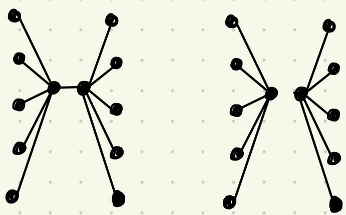
$$\left| q^*(G) - q^*(G \setminus E') \right| < \frac{2|E'|}{e(G)}$$

Robustness - OPEN

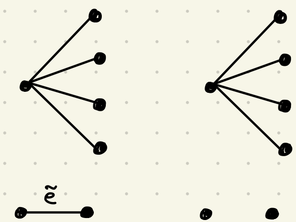
- <u>Robust</u> to small perturbations in edge set

$$\frac{-2}{e(G)} < q^*(G) - q^*(G\setminus\tilde{e}) < \frac{2}{e(G)}$$

m-edge G: Need $\frac{-1}{m}$ Need $\frac{2}{m}$



G $G\setminus\tilde{e}$



G $G\setminus\tilde{e}$

delete 1 edge

possible inc

$\frac{1}{m} \leq ? < \frac{2}{m}$

OPEN

possible dec

$\frac{2}{m}$

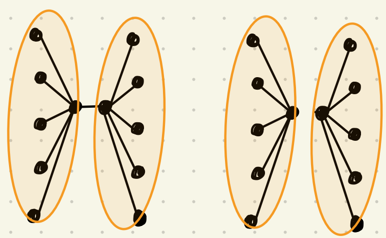
Robustness - OPEN

- Robust to small perturbations in edge set

$$\frac{-2}{e(G)} < q^*(G) - q^*(G\setminus\tilde{e}) < \frac{2}{e(G)}$$

m-edge G :

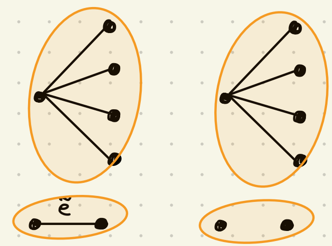
Need $\frac{-1}{m}$



G $\qquad$ $G\setminus\tilde{e}$

$q^* \sim \frac{1}{2} - \frac{1}{m}$ $\qquad$ $q^* = \frac{1}{2}$

Need $\frac{2}{m}$



G $\qquad$ $G\setminus\tilde{e}$

$q^* \sim \frac{2}{m}$ $\qquad$ $q^* = 0$

delete 1 edge

possible inc

$\frac{1}{m} \leqslant ? < \frac{2}{m}$

OPEN

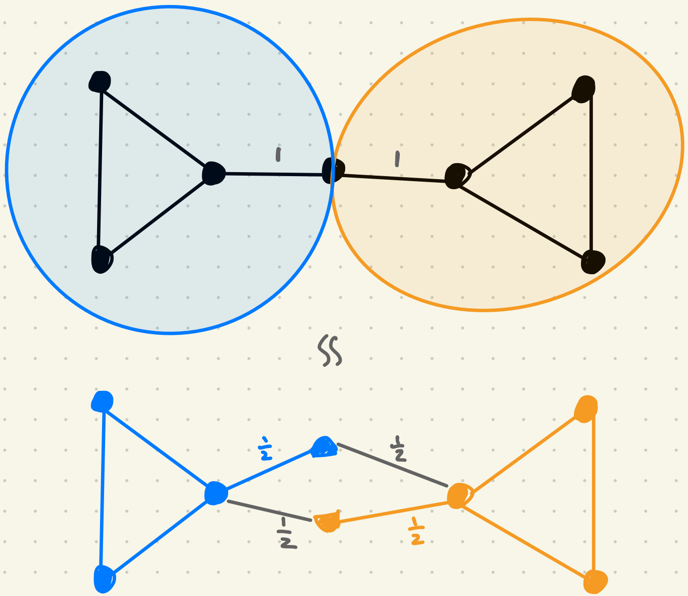
possible dec

$\frac{2}{m}$

Overlapping Communities - OPEN

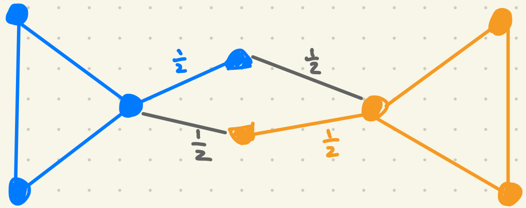
G, n vert. Q: #parts in OPT partion finite? $\leq 2^n$?

Usual Mod: Yes. $\leq n$ parts ($\leq \frac{n}{2}$ parts if no isol. vert)

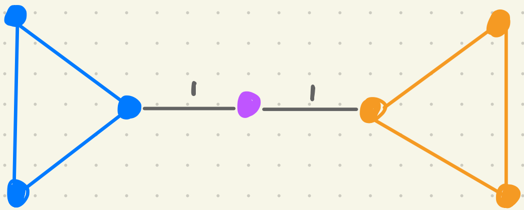


$$q = \frac{3\frac{1}{2} + 3\frac{1}{2}}{8} - \frac{7^2 + 7^2}{8^2} \approx 0.375$$

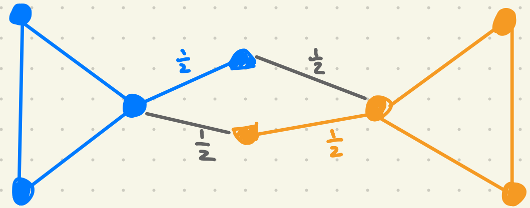
Overlapping Communities - OPEN



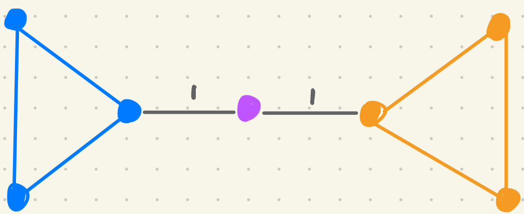
$$q = \frac{3\frac{1}{2} + 3\frac{1}{2}}{8} - \frac{8^2 + 8^2}{16^2} \approx 0.375$$



Overlapping Communities - OPEN

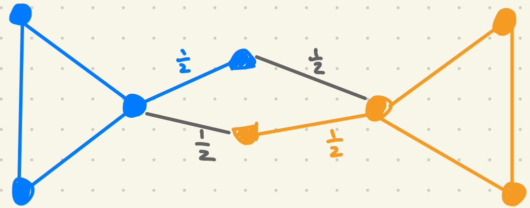


$$q = \frac{3\tfrac{1}{2} + 3\tfrac{1}{2}}{8} - \frac{8^2 + 8^2}{16^2} \approx 0.375$$

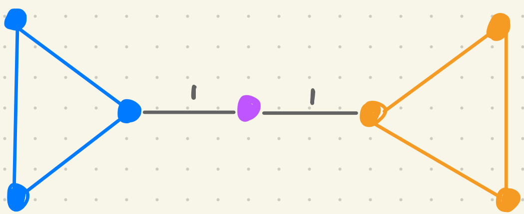


$$q = \frac{3 + 3}{8} - \frac{7^2 + 2^2 + 7^2}{16^2} \approx 0.353$$

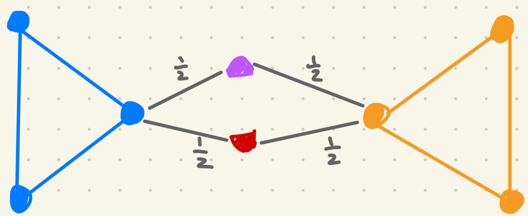
Overlapping Communities - Open



$$q = \frac{3\frac{1}{2} + 3\frac{1}{2}}{8} - \frac{8^2 + 8^2}{16^2} \approx 0.375$$

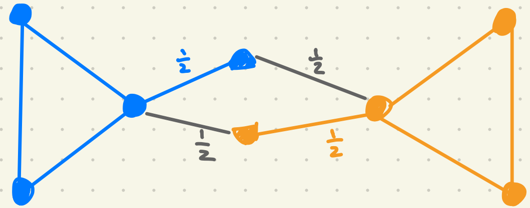


$$q = \frac{3 + 3}{8} - \frac{7^2 + 2^2 + 7^2}{16^2} \approx 0.353$$

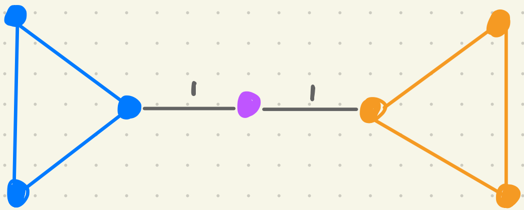


$$q = \frac{3 + 3}{8} - \frac{7^2 + 1^2 + 1^2 + 7^2}{16^2} \approx 0.360$$

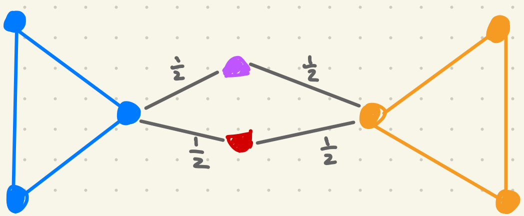
Overlapping Communities - OPEN



$$q = \frac{3\frac{1}{2} + 3\frac{1}{2}}{8} - \frac{8^2 + 8^2}{16^2} \approx 0.375$$



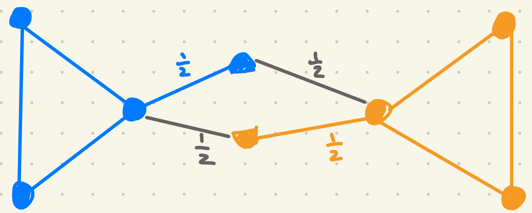
$$q = \frac{3 + 3}{8} - \frac{7^2 + 2^2 + 7^2}{16^2} \approx 0.353$$



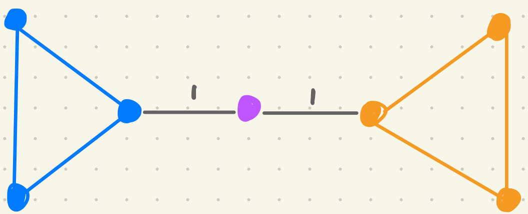
$$q = \frac{3 + 3}{8} - \frac{7^2 + 1^2 + 1^2 + 7^2}{16^2} \approx 0.360$$

$$q = \frac{3 + 3}{8} - \frac{7^2 + \frac{4}{k} + 7^2}{16^2} \approx 0.368$$

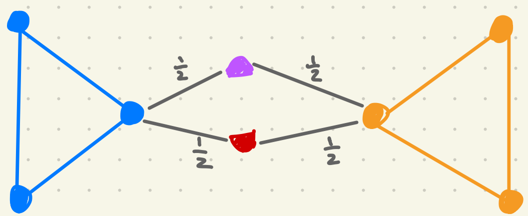
Overlapping Communities - OPEN



$$q = \frac{3\frac{1}{2} + 3\frac{1}{2}}{8} - \frac{8^2 + 8^2}{16^2} \approx 0.375$$



$$q = \frac{3 + 3}{8} - \frac{7^2 + 2^2 + 7^2}{16^2} \approx 0.353$$



$$q = \frac{3 + 3}{8} - \frac{7^2 + 1^2 + 1^2 + 7^2}{16^2} \approx 0.360$$

$$q = \frac{3 + 3}{8} - \frac{7^2 + \frac{4}{k} + 7^2}{16^2} \approx 0.368$$

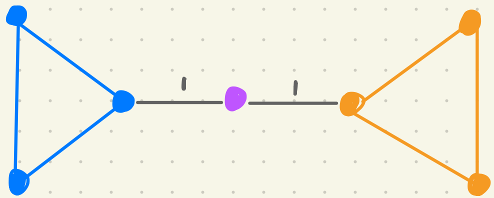
Overlapping Communities - OPEN

G, n vert. Q: #parts in OPT partion finite? $\leq 2^n$?

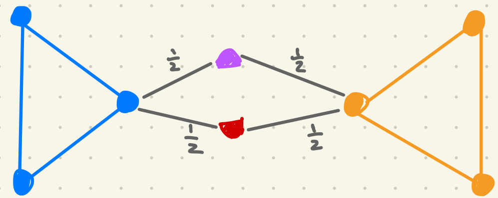
Usual Mod: Yes. $\leq n$ parts ($\leq \frac{n}{2}$ parts if no isol. vert)

Conj - for some graphs no finite #parts achieves OPT.

You can get "bad" parts to disappear ....



,



, ...

OPEN

Upper Bound for Modularity in terms of conductance h_G

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

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

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

Construction

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

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

Open

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

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

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

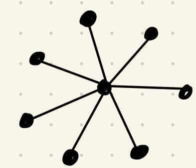
Recall

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

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

$$q^* \leq 1 - h_G$$

$\delta = 1$



$q^* = 0$

$\Rightarrow f(1) = 1$

$\delta \leq \varepsilon$

$q^* \geq 1 - \varepsilon$

$\Rightarrow f(0) = 0$