# Modularity of Erdős-Rényi random graphs 

Colin McDiarmid

Nice, July 2019
based on joint work with Fiona Skerman

## Modularity and communities

Modularity was introduced by Newman and Girvan in 2004 to give a measure of how well a graph can be divided into communities.

It now forms the backbone of the most popular algorithms used to cluster real data, with many applications, from protein discovery to identifying connections between websites.

See for example surveys by Fortunato (2010), and Porter Onnela and Mucha (2009), on the use of modularity for community detection in networks.

## Partitioning Networks:

## Network:

trade volume between countries
Garcia-Pérez 2016
USA, Canada, Bahamas, Haiti, Dominican Republic, Jamaica, Grenada, Mexico, Honduras, Venezuela, Peru

China, North Korea, Gambia, Sierra Leone, Togo, South Sudan

Japan, South Korea, Taiwan, Singapore, Sri Lanka, Philippines, New Zealand, Fiji, Papua New Guinea


## Definition of modularity

Let $G=(V, E)$ be a graph with $m \geq 1$ edges. For a set $A$ of vertices, let $e(A)$ be the number of edges within $A$, and let $\operatorname{vol}(A)$ be the sum over the vertices $v \in A$ of the degree $d_{v}$.

## Definition of modularity

Let $G=(V, E)$ be a graph with $m \geq 1$ edges. For a set $A$ of vertices, let $e(A)$ be the number of edges within $A$, and let $\operatorname{vol}(A)$ be the sum over the vertices $v \in A$ of the degree $d_{v}$.
Given a partition $\mathcal{A}$ of $V$, the modularity of $\mathcal{A}$ on $G$ is

$$
\begin{aligned}
q_{\mathcal{A}}(G) & =\frac{1}{2 m} \sum_{A \in \mathcal{A}} \sum_{u, v \in A}\left(\mathbf{1}_{u v \in E}-\frac{d_{u} d_{v}}{2 m}\right) \\
& =\frac{1}{m} \sum_{A \in \mathcal{A}} e(A)-\frac{1}{4 m^{2}} \sum_{A \in \mathcal{A}} \operatorname{vol}(A)^{2} ;
\end{aligned}
$$

and the modularity of $G$ is $q^{*}(G)=\max _{\mathcal{A}}(G)$.
Isolated vertices are irrelevant; and we shall not consider empty graphs (that is, with no edges).

## modularity: understanding the definition

$$
q_{\mathcal{A}}(G)=\frac{1}{2 m} \sum_{A \in \mathcal{A}} \sum_{u, v \in A}\left(\mathbf{1}_{u v \in E}-\frac{d_{u} d_{v}}{2 m}\right) .
$$

## modularity: understanding the definition

$$
q_{\mathcal{A}}(G)=\frac{1}{2 m} \sum_{A \in \mathcal{A}} \sum_{u, v \in A}\left(\mathbf{1}_{u v \in E}-\frac{d_{u} d_{v}}{2 m}\right) .
$$

If we pick uniformly at random a multigraph with the same degrees as $G$, then the expected number of edges between vertices $u$ and $v$ is essentially

$$
\frac{d_{u} d_{v}}{2 m}
$$

This is the rationale for the definition: whilst rewarding the partition for capturing edges within the parts, we should penalise by the expected number of edges.

## edge-contribution and degree tax

The second equation

$$
q_{\mathcal{A}}(G)=\frac{1}{m} \sum_{A \in \mathcal{A}} e(A)-\frac{1}{4 m^{2}} \sum_{A \in \mathcal{A}} \operatorname{vol}(A)^{2}
$$

expresses $q_{\mathcal{A}}(G)$ as the difference of two terms:
the edge contribution $q_{\mathcal{A}}^{E}(G)=\frac{1}{m} \sum_{A} e(A)$,
and the degree $\operatorname{tax} q_{\mathcal{A}}^{D}(G)=\frac{1}{4 m^{2}} \sum_{A} \operatorname{vol}(A)^{2}$.

## edge-contribution and degree tax

The second equation

$$
q_{\mathcal{A}}(G)=\frac{1}{m} \sum_{A \in \mathcal{A}} e(A)-\frac{1}{4 m^{2}} \sum_{A \in \mathcal{A}} \operatorname{vol}(A)^{2}
$$

expresses $q_{\mathcal{A}}(G)$ as the difference of two terms:
the edge contribution $q_{\mathcal{A}}^{E}(G)=\frac{1}{m} \sum_{A} e(A)$,
and the degree $\operatorname{tax} q_{\mathcal{A}}^{D}(G)=\frac{1}{4 m^{2}} \sum_{A} \operatorname{vol}(A)^{2}$.
Since $q_{\mathcal{A}}^{E}(G) \leq 1$ and $q_{\mathcal{A}}^{D}(G)>0$, we have $q_{\mathcal{A}}(G)<1$. Also, the trivial partition $\mathcal{A}_{0}$ with one part has $q_{\mathcal{A}_{0}}^{E}(G)=q_{\mathcal{A}_{0}}^{D}(G)=1$, so $q_{\mathcal{A}_{0}}(G)=0$. Thus

$$
0 \leq q^{*}(G)<1 .
$$

An example


## 3 possible partitions



$$
\begin{gathered}
q_{\mathcal{A}_{1}}^{E}=0.96, \quad q_{\mathcal{A}_{1}}^{D}=0.56 \\
q_{\mathcal{A}_{1}}=0.40
\end{gathered}
$$



$$
\begin{gathered}
q_{\mathcal{A}_{2}}^{E}=0.94, \quad q_{\mathcal{A}_{2}}^{D}=0.50 \\
q_{\mathcal{A}_{2}}=0.44
\end{gathered}
$$

$$
\begin{gathered}
q_{\mathcal{A}_{3}}^{E}=0.59, \quad q_{\mathcal{A}_{3}}^{D}=0.29 \\
q_{\mathcal{A}_{3}}=0.30
\end{gathered}
$$

## Modularity: some examples

(a) Let $G$ be a tree with $m$ edges and max degree $\Delta=o(m)$. Then $q^{*}(G)=1-o(1)$. (True also if treewidth $\Delta=o(m)$.)

## Modularity: some examples

(a) Let $G$ be a tree with $m$ edges and max degree $\Delta=o(m)$. Then $q^{*}(G)=1-o(1)$. (True also if treewidth $\Delta=o(m)$.)
(b) Let $G$ be an $m$-edge subgraph of the square lattice. Then $q^{*}(G)=1-o(1)$.

## Modularity: some examples

(a) Let $G$ be a tree with $m$ edges and max degree $\Delta=o(m)$. Then $q^{*}(G)=1-o(1)$. (True also if treewidth $\cdot \Delta=o(m)$.)
(b) Let $G$ be an m-edge subgraph of the square lattice. Then $q^{*}(G)=1-o(1)$.
(c) $q^{*}\left(K_{n}\right)=0$ (and indeed $q^{*}(G)=0$ if $G$ is $K_{n}$ less at most $n / 2$ edges).

## Modularity: some examples

(a) Let $G$ be a tree with $m$ edges and max degree $\Delta=o(m)$. Then $q^{*}(G)=1-o(1)$. (True also if treewidth $\cdot \Delta=o(m)$.)
(b) Let $G$ be an m-edge subgraph of the square lattice. Then $q^{*}(G)=1-o(1)$.
(c) $q^{*}\left(K_{n}\right)=0$ (and indeed $q^{*}(G)=0$ if $G$ is $K_{n}$ less at most $n / 2$ edges).

## Some properties of optimal partitions

Let $G$ have no isolated vertices, and let $\mathcal{A}$ be an optimal partition i.e. $q_{\mathcal{A}}(G)=q^{*}(G)$. Then each part $A$ in $\mathcal{A}$ induces a connected subgraph of $G$, with at least two vertices.
For example, if $G$ consists of $m$ disjoint edges, then the unique optimal partition has $m$ parts of size 2 , and $q^{*}(G)=1-1 / m$.

## Some properties of optimal partitions

Let $G$ have no isolated vertices, and let $\mathcal{A}$ be an optimal partition i.e. $q_{\mathcal{A}}(G)=q^{*}(G)$. Then each part $A$ in $\mathcal{A}$ induces a connected subgraph of $G$, with at least two vertices.
For example, if $G$ consists of $m$ disjoint edges, then the unique optimal partition has $m$ parts of size 2 , and $q^{*}(G)=1-1 / m$.

More generally, if $G$ consists of $k \geq 1$ cliques all of the same size, then

$$
q^{*}(G)=1-1 / k .
$$

## Resolution limit

Resolution limit (Fortunato and Barthélemy 2007).
Suppose that $G$ has $m$ edges and has a component $H$ with $<\sqrt{2 m}$ edges. Then $V(H)$ is a part in each optimal partition for $G$.

## Resolution Limit in pictures

Component $H$ $h$ edges


If $h<\sqrt{2 m}$, e.g. $m=1625$.


If $h>\sqrt{2 m}$, e.g. $m=1624$.
Graph $G, m$ edges

## Robustness

Optimal partition structure is sensitive to noise in edges.
The modularity value is robust:
if $G=(V, E)$ and $G^{\prime}=\left(V, E^{\prime}\right)$ are graphs with $|E| \geq\left|E^{\prime}\right|$, then

$$
\left|q^{*}(G)-q^{*}\left(G^{\prime}\right)\right| \leq \frac{2\left|E \backslash E^{\prime}\right|}{|E|}
$$

## Lusseau PhD Thesis


\# dolphins $=62$
\# edges $=159$
$q^{*}=0.52$

## Lusseau PhD Thesis


\# dolphins $=62$
\# edges $=159$
$q^{*}=0.52$
8.4\% of possible edges

## Random data


\# dolphins $=62$
say each pair interacts with probability 0.084
$q^{*}($ dolphins $)>q^{*}($ random network $) ? ?$

## Lusseau PhD Thesis


\# dolphins $=62$
\# edges $=159$
$q^{*}=0.52$
$8.4 \%$ of possible edges

## Random data

$q^{*}($ dolphins $)>q^{*}($ random network $) ? ?$

Modularity of Random Network on 62 vertices


Lusseau PhD Thesis

\# dolphins $=62$
\# edges $=159$
$q^{*}=0.52$
8.4\% of possible edges

## Random data

Simulate 62 vertices, with edge prob $p$.


Lusseau PhD Thesis

\# dolphins $=62$
\# edges $=159$
$q^{*}=0.52$
$8.4 \%$ of possible edges

## Random data

Simulate 1000 vertices, with edge prob $p$.


## Two theorems on $q^{*}(G(n, p)$

Here are two theorems of McD + Skerman on the modularity of random graphs $G(n, p)$. First, the overview.

Theorem (3 phases theorem)
(a) If $n^{2} p \rightarrow \infty$ and $n p \leq 1+o(1)$ then $q^{*}(G(n, p)) \xrightarrow{p} 1$.

## Two theorems on $q^{*}(G(n, p)$

Here are two theorems of McD + Skerman on the modularity of random graphs $G(n, p)$. First, the overview.

## Theorem (3 phases theorem)

(a) If $n^{2} p \rightarrow \infty$ and $n p \leq 1+o(1)$ then $q^{*}(G(n, p)) \xrightarrow{p} 1$.
(b) Given $1<c_{0} \leq c_{1}$, there exists $\delta>0$ such that, if

$$
c_{0} \leq n p \leq c_{1} \text {, then whp } \delta<q^{*}(G(n, p))<1-\delta .
$$

(c) If $n p \rightarrow \infty$ then $q^{*}(G(n, p)) \xrightarrow{p} 0$.

## Two theorems on $q^{*}(G(n, p)$

Here are two theorems of McD + Skerman on the modularity of random graphs $G(n, p)$. First, the overview.

## Theorem (3 phases theorem)

(a) If $n^{2} p \rightarrow \infty$ and $n p \leq 1+o(1)$ then $q^{*}(G(n, p)) \xrightarrow{p} 1$.
(b) Given $1<c_{0} \leq c_{1}$, there exists $\delta>0$ such that, if $c_{0} \leq n p \leq c_{1}$, then whp $\delta<q^{*}(G(n, p))<1-\delta$.
(c) If $n p \rightarrow \infty$ then $q^{*}(G(n, p)) \xrightarrow{p} 0$.

To prove part (a) it suffices to consider the partition into components. Part (c) and much of part (b) follow from the next theorem.

## Two theorems on $q^{*}(G(n, p)$

## Theorem (the ( $n p)^{-1 / 2}$ theorem)

There exist $0<a<b$ such that, if $n p \geq 1$ and $p \leq 0.99$, then

$$
\frac{a}{\sqrt{n p}}<q^{*}(G(n, p))<\frac{b}{\sqrt{n p}} \quad \text { whp. }
$$

This confirms a conjecture in 2006 by Reichardt and Bornholdt (and refutes another conjecture from the physics literature).

## Two theorems on $q^{*}(G(n, p)$

## Theorem (the ( $n p)^{-1 / 2}$ theorem)

There exist $0<a<b$ such that, if $n p \geq 1$ and $p \leq 0.99$, then

$$
\frac{a}{\sqrt{n p}}<q^{*}(G(n, p))<\frac{b}{\sqrt{n p}} \quad \text { whp. }
$$

This confirms a conjecture in 2006 by Reichardt and Bornholdt (and refutes another conjecture from the physics literature).

The upper bound may be proved using the expander mixing lemma (not here?).
The lower bound follows by considering a simple algorithm Swap (or, for a more limited range of $p$, from recent work on stochastic block models.)

## Two theorems on $q^{*}(G(n, p)$

As we noted, much of the $n p>1$ part of the 3 phases theorem follows from the $(n p)^{-1 / 2}$ theorem.

To complete the proof for $n p>1$, we need to show that $q^{*}(G(n, p))<1-\delta$ whp when $n p$ is just above 1 .

To do this, we may use the result that whp, splitting the giant component roughly into halves must break $\Omega(n)$ edges (Luczak and McD 2001).

## Swap gives the $(n p)^{-1 / 2}$ lower bound

Given a graph $G$, the algorithm Swap runs in linear time and yields a balanced bipartition $\mathcal{A}$ of the vertices.

## Theorem

There are constants $c_{0}$ and $a>0$ such that if $p=p(n)$ satisfies $1 \leq n p \leq n-c_{0}$, then whp

$$
q_{\mathcal{A}}\left(G_{n, p}\right) \geq a\left(\frac{1-p}{n p}\right)^{1 / 2}
$$

and if also $n p \geq c_{0}$ we may take $a=\frac{1}{5}$.

## Idea of Swap

The algorithm Swap starts with a balanced bipartition of the vertex set into $A \cup B$, which has modularity very near 0 whp.

By swapping some pairs between $A$ and $B$, whp we can increase the edge contribution significantly, without changing the distribution of the degree tax (and without introducing dependencies which would be hard to analyse).

## The algorithm Swap

Assume for simplicity that $6 \mid n$ and write $n=6 k$. Start with the bipartition $\mathcal{A}$ of $V=[n]$ into $A=\{j \in V: j$ is odd $\}$ and $B=\{j \in V: j$ is even $\}$. Whp $q_{\mathcal{A}}\left(G_{n, p}\right)$ is very close to 0 .

## The algorithm Swap

Assume for simplicity that $6 \mid n$ and write $n=6 k$. Start with the bipartition $\mathcal{A}$ of $V=[n]$ into $A=\{j \in V: j$ is odd $\}$ and $B=\{j \in V: j$ is even $\}$. Whp $q_{\mathcal{A}}\left(G_{n, p}\right)$ is very close to 0 . Let $V_{0}=[4 k]$, let $V_{1}=\{4 k+1, \ldots, 6 k\}$. Let $A_{0}=A \cap V_{0}$, $A_{1}=A \cap V_{1}$ and $B_{0}=B \cap V_{0}, B_{1}=B \cap V_{1}$. The four sets $A_{i}, B_{i}$ are pairwise disjoint with union $V$.

## The algorithm Swap

Assume for simplicity that $6 \mid n$ and write $n=6 k$. Start with the bipartition $\mathcal{A}$ of $V=[n]$ into $A=\{j \in V: j$ is odd $\}$ and $B=\{j \in V: j$ is even $\}$. Whp $q_{\mathcal{A}}\left(G_{n, p}\right)$ is very close to 0 .
Let $V_{0}=[4 k]$, let $V_{1}=\{4 k+1, \ldots, 6 k\}$. Let $A_{0}=A \cap V_{0}$,
$A_{1}=A \cap V_{1}$ and $B_{0}=B \cap V_{0}, B_{1}=B \cap V_{1}$. The four sets $A_{i}, B_{i}$ are pairwise disjoint with union $V$.

Initially $V_{0}$ is partitioned into $A_{0} \cup B_{0}$ : the algorithm Swap 'improves' this partition, keeping $A_{1}, B_{1}$ fixed. For $i=1, \ldots, 2 k$ let $a_{i}=2 i-1$ and $b_{i}=2 i$, so $A_{0}=\left\{a_{1}, \ldots, a_{2 k}\right\}$ and $B_{0}=\left\{b_{1}, \ldots, b_{2 k}\right\}$. We improve the partition $V_{0}=A_{0} \cup B_{0}$ is by independently swapping $a_{i}$ and $b_{i}$ for certain values $i$.

## $T_{i}$ and swapping $a_{i}, b_{i}$

For each $i \in[2 k]$ let

$$
T_{i}=e\left(a_{i}, B_{1}\right)-e\left(a_{i}, A_{1}\right)+e\left(b_{i}, A_{1}\right)-e\left(b_{i}, B_{1}\right)
$$

The random variables $T_{1}, \ldots, T_{2 k}$ are iid.
Also $\mathbb{E}\left[T_{i}\right]=0, \operatorname{var}\left(T_{i}\right)=4 \mathrm{kp}(1-\mathrm{p})$; and $\mathbb{E}\left[\left|T_{i}\right|\right]=\Theta\left((n p(1-p))^{1 / 2}\right)$.

## $T_{i}$ and swapping $a_{i}, b_{i}$

For each $i \in[2 k]$ let

$$
T_{i}=e\left(a_{i}, B_{1}\right)-e\left(a_{i}, A_{1}\right)+e\left(b_{i}, A_{1}\right)-e\left(b_{i}, B_{1}\right)
$$

The random variables $T_{1}, \ldots, T_{2 k}$ are iid.
Also $\mathbb{E}\left[T_{i}\right]=0, \operatorname{var}\left(\mathrm{~T}_{\mathrm{i}}\right)=4 \mathrm{kp}(1-\mathrm{p})$; and $\mathbb{E}\left[\left|T_{i}\right|\right]=\Theta\left((n p(1-p))^{1 / 2}\right)$.
If $T_{i}>0$ and we swap $a_{i}$ and $b_{i}$ between $A_{0}$ and $B_{0}$ (that is, replace $A_{0}$ by $\left(A_{0} \backslash\left\{a_{i}\right\}\right) \cup\left\{b_{i}\right\}$ and similarly for $\left.B_{0}\right)$ then $e(A, B)$ decreases by $T_{i}$, so the edge contribution of the partition increases.

## Illustration of swapping



## $T^{*}$ and swaps

Swap makes all such swaps (looking only at possible edges between $V_{0}$ and $\left.V_{1}\right)$, yielding the balanced bipartition $\mathcal{A}^{\prime}=\left(A^{\prime}, B^{\prime}\right)$, where $A^{\prime}=A_{0}^{\prime} \cup A_{1}$ and $B^{\prime}=B_{0}^{\prime} \cup B_{1}$.
Let $T^{*}=\sum_{i \in[2 k]}\left|T_{i}\right|$. Then

$$
e\left(A_{0}^{\prime}, A_{1}\right)+e\left(B_{0}^{\prime}, B_{1}\right)-\left(e\left(A_{0}^{\prime}, B_{1}\right)+e\left(A_{1}, B_{0}^{\prime}\right)\right)=T^{*} .
$$

## $T^{*}$ and swaps

Swap makes all such swaps (looking only at possible edges between $V_{0}$ and $\left.V_{1}\right)$, yielding the balanced bipartition $\mathcal{A}^{\prime}=\left(A^{\prime}, B^{\prime}\right)$, where $A^{\prime}=A_{0}^{\prime} \cup A_{1}$ and $B^{\prime}=B_{0}^{\prime} \cup B_{1}$.

Let $T^{*}=\sum_{i \in[2 k]}\left|T_{i}\right|$. Then

$$
e\left(A_{0}^{\prime}, A_{1}\right)+e\left(B_{0}^{\prime}, B_{1}\right)-\left(e\left(A_{0}^{\prime}, B_{1}\right)+e\left(A_{1}, B_{0}^{\prime}\right)\right)=T^{*} .
$$

But $e\left(A_{0}^{\prime}, A_{1}\right)+e\left(B_{0}^{\prime}, B_{1}\right)+\left(e\left(A_{0}^{\prime}, B_{1}\right)+e\left(A_{1}, B_{0}^{\prime}\right)\right)=e\left(V_{0}, V_{1}\right)$, so

$$
e\left(A_{0}^{\prime}, A_{1}\right)+e\left(B_{0}^{\prime}, B_{1}\right)=\frac{1}{2} e\left(V_{0}, V_{1}\right)+\frac{1}{2} T^{*} .
$$

## $T^{*}$ and edge contribution

$T^{*}$ is the sum of the $2 k \approx n / 3$ iid random variables $\left|T_{i}\right|$, so whp

$$
T^{*} \approx 2 k \mathbb{E}\left[\left|T_{1}\right|\right]=\Theta\left(n^{3 / 2}(p(1-p))^{1 / 2}\right)
$$

## $T^{*}$ and edge contribution

$T^{*}$ is the sum of the $2 k \approx n / 3$ iid random variables $\left|T_{i}\right|$, so whp

$$
T^{*} \approx 2 k \mathbb{E}\left[\left|T_{1}\right|\right]=\Theta\left(n^{3 / 2}(p(1-p))^{1 / 2}\right)
$$

Thus whp the edge contribution for $\mathcal{A}^{\prime}$ beats that for the initial bipartition $\mathcal{A}$ by

$$
\Theta\left(\frac{n^{3 / 2}(p(1-p))^{1 / 2}}{n^{2} p}\right)=\Theta\left(\left(\frac{1-p}{n p}\right)^{1 / 2}\right) .
$$

In other words

$$
q_{\mathcal{A}^{\prime}}^{E}\left(G_{n, p}\right)-q_{\mathcal{A}}^{E}\left(G_{n, p}\right)=\Theta\left(\left(\frac{1-p}{n p}\right)^{1 / 2}\right) \quad \text { whp. }
$$

## What about degree tax?

Our decisions about when to swap are symmetric. In the two cases

$$
\begin{aligned}
& e\left(a_{i}, B_{1}\right)=w, e\left(a_{i}, A_{1}\right)=x \text { and } e\left(b_{i}, A_{1}\right)=y, e\left(b_{i}, B_{1}\right)=z \\
& e\left(a_{i}, B_{1}\right)=y, e\left(a_{i}, A_{1}\right)=z \text { and } e\left(b_{i}, A_{1}\right)=w, e\left(b_{i}, B_{1}\right)=x .
\end{aligned}
$$

we make the same decision (swap iff $w-x+y-z>0$ ). It follows that the degree tax for $\mathcal{A}^{\prime}$ has exactly the same distribution as for $\mathcal{A}$. We find

$$
q_{\mathcal{A}^{\prime}}^{D}\left(G_{n, p}\right)-q_{\mathcal{A}}^{D}\left(G_{n, p}\right)=o\left(\left(\frac{1-p}{n p}\right)^{1 / 2}\right) \text { whp. }
$$

## completing the Swap story

Putting together the results on edge contribution and on degree tax we find

$$
q_{\mathcal{A}^{\prime}}\left(G_{n, p}\right)-q_{\mathcal{A}}\left(G_{n, p}\right)=\Theta\left(\left(\frac{1-p}{n p}\right)^{1 / 2}\right) \text { whp. }
$$

But whp $q_{\mathcal{A}}\left(G_{n, p}\right)$ is very near 0 , and so

$$
q_{\mathcal{A}^{\prime}}\left(G_{n, p}\right)=\Theta\left(\left(\frac{1-p}{n p}\right)^{1 / 2}\right) \quad \text { whp }
$$

as required.

## a reference, including further references

Colin McDiarmid and Fiona Skerman, Modularity of Erdős-Rényi random graphs, LIPIcs 29th AofA 2018: Uppsala, Sweden, June 2018, 31:1 31:13.

## a reference, including further references

Colin McDiarmid and Fiona Skerman, Modularity of Erdős-Rényi random graphs, LIPIcs 29th AofA 2018: Uppsala, Sweden, June 2018, 31:1 31:13.

