Misleading stars: What cannot be measured on the Internet?
Map of Online Communities and Related Points of Interest
Geographic area represents estimated size of membership.

(Not a complete survey. Sizes based on best figures I could find but involved some guesswork. Do not use for navigation.)
How accurate are network maps?

Why?
- To develop/adapt protocols to Internet PaDIS, RMTP
- To understand the impact of uncertainty: networks metrology?

How?
- multicast [Marchetta et al., JSAC’11]
- network tomography
- Traceroute

Takeaway
- Compare internet topologies instead of counting them
- Local properties suffer
- Global properties suffer less
Idea: send "buggy" TTL packets, collect error messages
Traceroute Limitations

- **Sampling bias**: Impact of sources location
- **Aliasing**: How to map IPs to the equipment
- **Load-Balancing**: Traceroute assumes all packets follow the same path...
- **Stars**: Disabled/Filtered ICMP messages

![Diagram](image)

Trace:

\[ t_1 = a, *, 1, b \]
\[ t_2 = b, *, 1, i_1, i_2, i_3, *, 3, c \]
**Sampling bias**: Impact of sources location

**Aliasing**: How to map IPs to the equipment

**Load-Balancing**: Traceroute assumes all packets follow the same path...

**Stars**: Disabled/Filtered ICMP messages

\[
\begin{align*}
\text{Trace:} & \quad t_1 = a, \star_1, b \\
& \quad t_2 = b, \star_2, i_1, i_2, i_3, \star_3, c
\end{align*}
\]
Traceroute Limitations

- **Sampling bias**: Impact of sources location
- **Aliasing**: How to map IPs to the equipment
- **Load-Balancing**: Traceroute assumes all packets follow the same path...
- **Stars**: Disabled/Filtered ICMP messages

Trace:

\[ t_1 = a, \star_1, b \]
\[ t_2 = b, \star_2, i_1, i_2, i_3, \star_3, c \]

?
Traceroute Limitations

- **Sampling bias**: Impact of sources location
- **Aliasing**: How to map IPs to the equipment
- **Load-Balancing**: Traceroute assumes all packets follow the same path...
- **Stars**: Disabled/Filtered ICMP messages

Trace:

\[ t_1 = a, \star_1, b \]
\[ t_2 = b, \star_2, i_1, i_2, i_3, \star_3, c \]
• **Sampling bias**: Impact of sources location
• **Aliasing**: How to map IPs to the equipment
• **Load-Balancing**: Traceroute assumes all packets follow the same path...
• **Stars**: Disabled/Filtered ICMP messages

Trace:

\[ t_1 = a, \ast_1, b \]
\[ t_2 = b, \ast_2, i_1, i_2, i_3, \ast_3, c \]
Traceroute Limitations

- **Sampling bias**: Impact of sources location
- **Aliasing**: How to map IPs to the equipment
- **Load-Balancing**: Traceroute assumes all packets follow the same path...
- **Stars**: Disabled/Filtered ICMP messages

Trace:

\[ t_1 = a, \star_1, b \]
\[ t_2 = b, \star_2, i_1, i_2, i_3, \star_3, c \]
Related Works

Works initiated by Acharya and Gouda [SSS09, ICDCN10, ICDCN11]

- Models to capture *irregular* nodes
- Irregular nodes $\rightarrow$ anonymous nodes
- Central concept of *minimal* topologies
- Counting the number of generable topologies

Our approach: Lots of generable topologies is not a problem if they are all *similar*
- $G_0(V_0, E_0)$: static undirected graph = **Target Topology**.

- $v \in V_0$ is
  - either named: always answers with its only name (no aliasing)
  - either anonymous: always answers $\star$.

- Not necessarily minimal!

- $d_{G_0}(u, v)$ is the shortest path distance.
Model

- $G_0(V_0, E_0)$: static undirected graph = Target Topology.
- $v \in V_0$ is
  - either named: always answers with its only name (no aliasing)
  - either anonymous: always answers ⋆.
- Not necessarily minimal!
- $d_{G_0}(u, v)$ is the shortest path distance.
We don’t know $G_0$, but we know a set of traces $\mathcal{T}$ of $G_0$.

- Anonymous nodes appear as stars ($\star$) in $\mathcal{T}$
- Each star has a unique number ($\star_i$, $i = 1..s$) in $\mathcal{T}$
- $d_T(u, v)$ is the number of symbols in $T \in \mathcal{T}$ between $u$ and $v$.
- No Assumption on the number of traces, on path uniqueness nor symmetry.

Anonymous nodes appear as stars ($\star$) in $\mathcal{T}$

Each star has a unique number ($\star_i$, $i = 1..s$) in $\mathcal{T}$

$d_T(u, v)$ is the number of symbols in $T \in \mathcal{T}$ between $u$ and $v$.

No Assumption on the number of traces, on path uniqueness nor symmetry.

\[ t_1 = a, \star_1, b \]
\[ t_2 = b, \star_2, i_1, i_2, i_3, \star_3, c \]
\[ \leftrightarrow \]

\[ t_1 = b, \star_1, c \]
\[ t_2 = b, \star_2, c \]
\[ t_3 = a, b, \star_3, c \]
\[ t_4 = c, \star_4, b, a \]
**Complete cover:** Each edge of $G_0$ appears at least once in some trace to $\mathcal{T}$

**Reality sampling:** For every trace $T \in \mathcal{T}$, if the distance between two symbols $\sigma_1, \sigma_2 \in T$ is $d_T(\sigma_1, \sigma_2) = k$, then there exists a path (i.e., a walk without cycles) of length $k$ connecting two (named or anonymous) nodes $\sigma_1$ and $\sigma_2$ in $G_0$.

No assumption on coverage

\[ a, b, c \]
\[ a, c, b \]
\[ a, *_1, c \]
\[ a, d, c \]
\(\alpha\)-(Routing) Consistency: There exists an \(\alpha \in (0, 1]\) such that, for every trace \(T \in \mathcal{T}\), if \(d_T(\sigma_1, \sigma_2) = k\) for two entries \(\sigma_1, \sigma_2\) in trace \(T\), then the shortest path connecting the two (named or anonymous) nodes corresponding to \(\sigma_1\) and \(\sigma_2\) in \(G_0\) has distance at least \(\lceil \alpha k \rceil\).

Note: \(\alpha > 0 \iff \text{loop-less routing}\)

\[a, b, c \quad a, i_1, i_3, i_4, c \Rightarrow \alpha \leq \frac{2}{5}\]
A topology $G$ is $\alpha$-consistently \textbf{inferrable} from $\mathcal{T}$ if it respects the 3 previous rules.

Let $\mathcal{G}_T = \{G, \text{ s.t. } G \text{ is inferrable from } \mathcal{T}\}$. We study the properties of the set $\mathcal{G}_T$ of inferrable topologies.

We define the \textbf{canonic graph} $G_c$ as the straightforward graph that treats each star as unique.
1 inferrable topology = 1 mapping of stars to anonymous routers.

Let Map be such function.
G_c is inferrable. Map = Id.

(i) if \( \star_1 \in T_1 \) and \( \star_2 \in T_2 \), and
\[
\left[ \alpha \cdot d_{T_1}(\star_1, u) \right] > d_C(u, \star_2) \Rightarrow \text{Map}(\star_1) \neq \text{Map}(\star_2).
\]

(ii) \( \star_1 \in T_1 \) \( \star_2 \in T_2 \), and \( \exists T \) s.t.
\[
\left[ \alpha \cdot d_T(u, v) \right] > d_C(u, \star_1) + d_C(v, \star_2) \Rightarrow \text{Map}(\star_1) \neq \text{Map}(\star_2).
\]
Algorithm constructive part

We construct the Star Graph \( G_*(V_*, E_*) \):
- Vertices=stars in the trace
- Edges=if stars cannot be merged:
  \((\star_1, \star_2) \in E_* \iff \operatorname{Map}(\star_1) \neq \operatorname{Map}(\star_2)\).

1 proper coloring of \( G_* \) ↔ 1 Map function

- minimal coloring \(\rightarrow\) minimal topology
- ’maximal’ coloring \(\rightarrow\) \( G_c \)

\[
\sum_{k=\gamma(G_*)}^{\vert V_* \vert} \frac{P(G_*, k)}{k!} \geq \vert G_T \vert, \\
\gamma(G_*) = \text{chromatic number of } G_* \\
P(G_*, k) = \text{chromatic polynomial of } G_*.
\]
Trace:
\[ t_1 = a, \star_1, b \]
\[ t_2 = b, \star_2, i_1, i_2, i_3, \star_3, c \]
\[ \alpha = 0.5 \]
Results are **bad**!

- **Connected components**
  No assumption on "coverage" \(\Rightarrow\)
  Stars disconnect the graph!
  \[
  |cc(G_1) / cc(G_2)| \leq \frac{n}{2}
  \]

- **Stretch**
  Even if we only consider connected topologies.
  \[
  |stretch(G)| \leq \frac{n+s-1}{2}
  \]
It's worse!

- **Triangles**
  Bipartite complete graph worst case
  \[ |C_3(G_1)/C_3(G_2)| \leq \infty \]

- **Degree**
  Worst case is on anonymous nodes
  \[ |\text{DEG}(G_1) - \text{DEG}(G_2)| \leq 2(s - \gamma(G_\star)) \]
Best case: Fully explored topologies

Strong assumptions:

- $\alpha = 1$ : shortest path routing
- $\forall u, v \in V_0, \exists T \in T$ such that $u \in T \lor v \in T$
- We don’t see any stronger case

Results:

- Global properties conserved
- Local properties: still bad!

Yvonne-Anne Pignolet, Stefan Schmid, G. Trédan

Misleading stars
## Overall Results

### Absolute difference: $G_1 - G_2$

<table>
<thead>
<tr>
<th>Property</th>
<th>Arbitrary</th>
<th>Fully Explored ($\alpha = 1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td># of nodes</td>
<td>$\leq s - \gamma(G_*)$</td>
<td>$\leq s - \gamma(G_*)$</td>
</tr>
<tr>
<td># of links</td>
<td>$\leq 2(s - \gamma(G_*)$</td>
<td>$\leq 2(s - \gamma(G_*)$</td>
</tr>
<tr>
<td># of CC</td>
<td>$\leq n/2$</td>
<td>$= 0$</td>
</tr>
<tr>
<td>Diameter</td>
<td>$\leq (s - 1)/s \cdot (N - 1)$</td>
<td>$s/2$ (¶)</td>
</tr>
<tr>
<td>Max. Deg.</td>
<td>$\leq 2(s - \gamma(G_*)$</td>
<td>$\leq 2(s - \gamma(G_*)$</td>
</tr>
<tr>
<td>Triangles</td>
<td>$\leq 2s(s - 1)$</td>
<td>$\leq 2s(s - 1)/2$</td>
</tr>
</tbody>
</table>

### Relative difference: $G_1 / G_2$

<table>
<thead>
<tr>
<th>Property</th>
<th>Arbitrary</th>
<th>Fully Explored ($\alpha = 1$)</th>
</tr>
</thead>
<tbody>
<tr>
<td># of nodes</td>
<td>$\leq (n + s)/(n + \gamma(G_*)$</td>
<td>$\leq (n + s)/(n + \gamma(G_*)$</td>
</tr>
<tr>
<td># of links</td>
<td>$\leq (\nu + 2s)/(\nu + 2)$</td>
<td>$\leq (\nu + 2s)/(\nu + 2)$</td>
</tr>
<tr>
<td># of CC</td>
<td>$\leq n/2$</td>
<td>$= 1$</td>
</tr>
<tr>
<td>Stretch</td>
<td>$\leq (N - 1)/2$</td>
<td>$= 1$</td>
</tr>
<tr>
<td>Diameter</td>
<td>$\leq s$</td>
<td>$2$</td>
</tr>
<tr>
<td>Max. Deg.</td>
<td>$\leq s - \gamma(G_*) + 1$</td>
<td>$\leq s - \gamma(G_*) + 1$</td>
</tr>
<tr>
<td>Triangles</td>
<td>$\infty$</td>
<td>$\infty$</td>
</tr>
</tbody>
</table>
Conclusion

- Constructive proofs
- Complex algorithm
- Don’t count, compare!
- Huge dissimilarities
- Worst case approach

A practical part:
- compare with reality
- develop property estimation algorithms
Thanks!