Subgraphs, Motifs, and Scaling in a Dynamic Airline Network

arXiv:1807.02585

Steve Lawford

ENAC, University of Toulouse

Online, 26 August 2022
(Econophysics Colloquium)

Joint work with Marius Agasse-Duval

Introduction

- Small-scale topological structure of dynamic airline network
- Graph-theoretic research on transportation networks typically focuses on macro or micro measures (diameter, centrality)
- Little known about subgraph-level behaviour (scaling, motifs), sometimes called "mesoscopic" measures

(1) **Count small subgraphs** — exact analytic enumeration

(2) **Investigate scaling** — power law between subgraph counts and edges, evidence for model evolution ("phase transition")

(3) **Identify motifs** — statistically significant wrt. random graphs

Framework for study of simple undirected graphs

- Real-world network as graph

$$G = (V, E), \quad n = |V|, \quad m = |E|$$

- Adjacency matrix

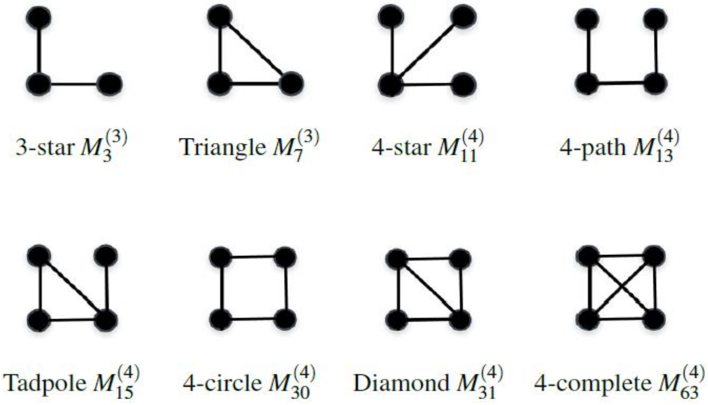
$$g = (g)_{ij}, \quad (g)_{ii} = 0, \quad (g)_{ij} = (g)_{ji}, \quad (g)_{ij} \in \{0, 1\}$$

- Edge, neighbourhood, degree, density

$$(i, j) \in E, \quad \Gamma_G(i) = \{j : (i, j) \in E\}$$

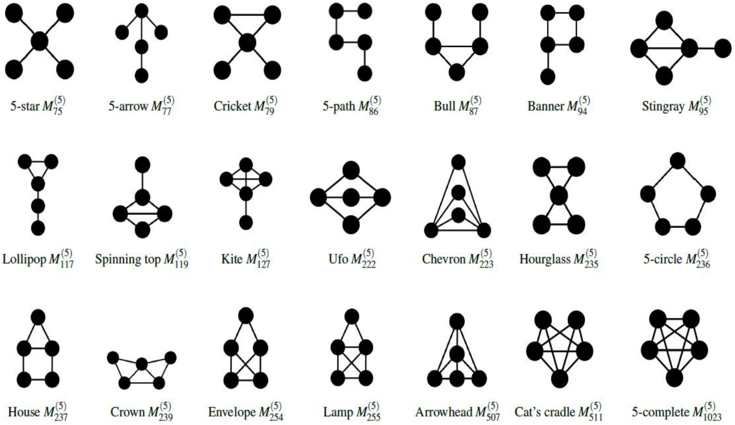
$$k_i = |\Gamma_G(i)|, \quad d(G) = 2m/n(n-1)$$

Small connected non-isomorphic subgraphs (3, 4 nodes)



Subgraph $M_a^{(b)}$ is $G' = (V', E') \subseteq G$ with $V' \subseteq V$ and $E' \subseteq E$, such that $(i,j) \in E' \implies i,j \in V'$ (induced subgraph $\widetilde{M}_a^{(b)}$ has every possible edge)

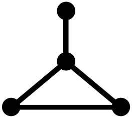
Small connected non-isomorphic subgraphs (5 nodes)



e.g. $\mathrm{M}_a^{(b)}$, $b = 5$, $a = \sum_{i=1}^{b-1} \sum_{j=i+1}^{b} 2^{\binom{b-i}{2}+(b-j)} (g)_{ij}$, $119_{10} = 0001110111_2$ (spinning top)

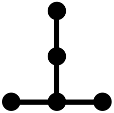
Counting subgraphs by decomposition

Let g be the adjacency matrix and let k_i be the degree of node i



tadpole

$$|M_{15}^{(4)}| = \frac{1}{2} \sum_{k_i > 2} (g^3)_{ii} \, (k_i - 2)$$

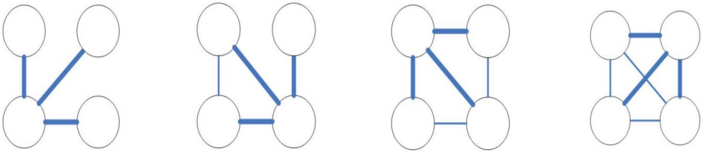


5-arrow

$$|M_{77}^{(5)}| = \sum_{\substack{(i,j) \in E \\ \text{both directions}}} \binom{k_i - 1}{2} (k_j - 1) - 2 |M_{15}^{(4)}|$$

Generally, analytic formulae for subgraphs can be expressed in terms of simpler subgraphs (Noga Alon et al., 1997; Estrada, 2011)

Counting induced subgraphs by linear combination



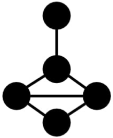
(a) *4-star* $M_{11}^{(4)}$. **(b)** *Tadpole* $M_{15}^{(4)}$. **(c)** *Diamond* $M_{31}^{(4)}$. **(d)** *4-complete* $M_{63}^{(4)}$.

$$|\widetilde{M}_{11}^{(4)}| = |M_{11}^{(4)}| - |\widetilde{M}_{15}^{(4)}| - 2\,|\widetilde{M}_{31}^{(4)}| - 4\,|M_{63}^{(4)}|$$
$$|\widetilde{M}_{15}^{(4)}| = |M_{15}^{(4)}| - 4\,|\widetilde{M}_{31}^{(4)}| - 12\,|M_{63}^{(4)}|$$
$$|\widetilde{M}_{31}^{(4)}| = |M_{31}^{(4)}| - 6\,|M_{63}^{(4)}|$$

$$\Longrightarrow\ |\widetilde{M}_{11}^{(4)}| = |M_{11}^{(4)}| - |M_{15}^{(4)}| + 2\,|M_{31}^{(4)}| - 4\,|M_{63}^{(4)}|$$

Exact enumeration much harder for larger subgraphs

Let g be the adjacency matrix and k_i the degree of node i, and $S(i,j) = \Gamma_G(i) \cap \Gamma_G(j)$ the common neighbourhood of nodes i, j



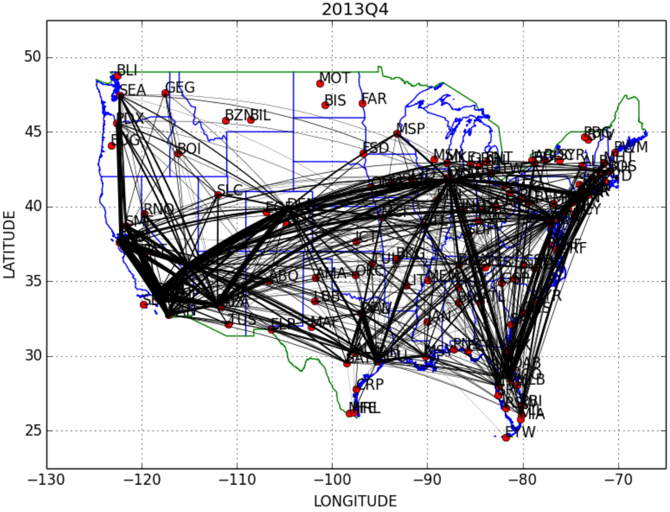
spinning top

$$|M_{119}^{(5)}| = \sum_{\substack{(i,j)\in E \\ k_i>2,\, k_j>2}} ((g^2)_{ij}-1) \sum_{\substack{r\in S(i,j) \\ k_r>1}} (k_r-2) - 12\,|M_{63}^{(4)}|$$

$$|\widetilde{M}_{119}^{(5)}| = |M_{119}^{(5)}| - 3\,|M_{127}^{(5)}| - 2\,|M_{239}^{(5)}| - 2\,|M_{254}^{(5)}| + 8\,|M_{255}^{(5)}| + 8\,|M_{507}^{(5)}| - 24\,|M_{511}^{(5)}| + 60\,|M_{1023}^{(5)}|$$

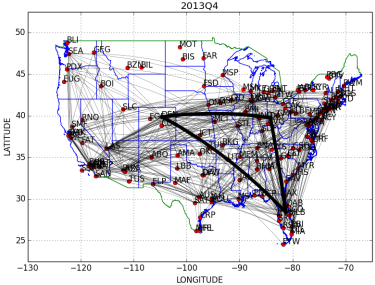
Much work on efficient algorithms to make subgraph counting feasible for larger subgraphs and networks (Ribeiro et al., 2021)

Route-map data for Southwest Airlines (1)

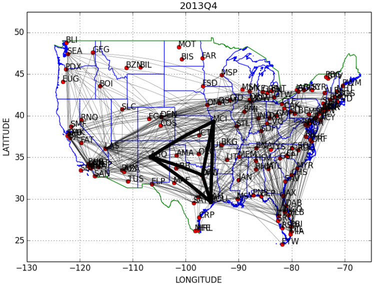


DB1B, 1999Q1 – 2013Q4, unidirectional route-level, direct tickets

Route-map data for Southwest Airlines (2)[1]



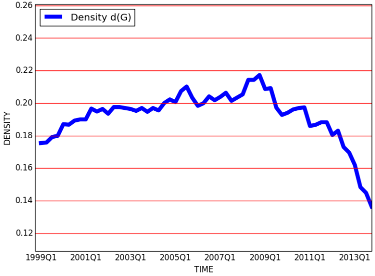
Triangle $M_7^{(3)}$



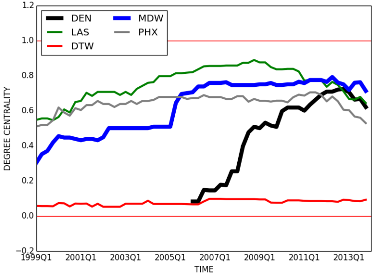
4-complete $M_{63}^{(4)}$

[1] Dayton–Denver–Orlando and Albuquerque–Dallas–Houston–Kansas City

Network dynamics — local and global



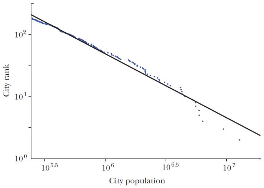
Density



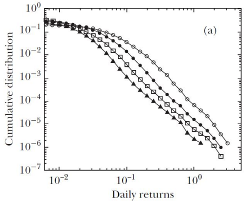
Degree centrality

- Number of airports has increased much faster than routes
- Heterogeneity in degree centrality over time, different airports
- Diameter and average path length very similar to Erdős-Rényi $G(n, p)$ but much higher clustering, suggesting small world, and subgraph counts generally much higher than in $G(n, p)$

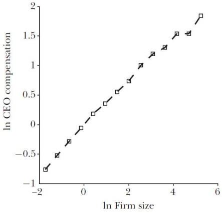
Power laws in economics and biology (Gabaix, 2016)[2]



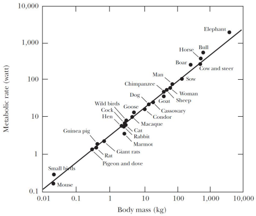
$\beta \approx -1$ (Zipf's law)



$\beta \approx -3$ (cubic law)



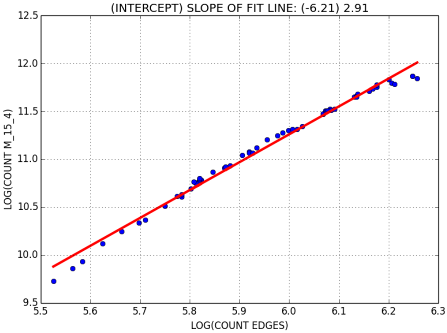
$\beta \approx 1/3$ (Robert's law)



$\beta \approx 3/4$ (Kleiber's law)

[2] Clockwise from top-left: city rank and population, stock return distribution and return, metabolic rate and body mass, CEO compensation and firm size

How do subgraphs scale with network size?[3]

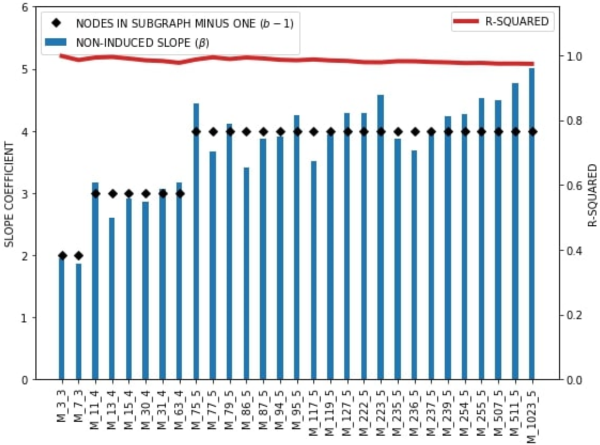


tadpole

- Evidence that $|M_a^{(b)}| = A\, m^{\beta}$ with $\beta \approx b - 1$
- Erdős-Rényi $G(n, p)$, general scaling result (Bollobás, 1985; Itzkovitz & Uri Alon, 2005) $\implies$ $\beta = b/2$ as n large

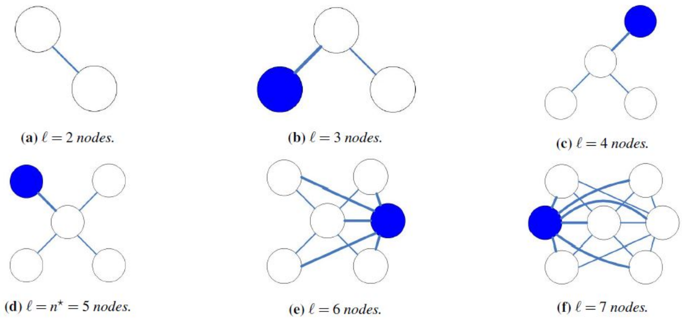
[3] "Networks are full of power laws . . ." but do they arise from optimality or randomness? "It would be nice to know." (Gabaix, 2016, pp.200–201)

Evidence for power law $|M_a^{(b)}| = A\, m^{\beta}$, full-sample



- Close match between estimated slope β (blue), and $b-1$ (black) for *all* subgraphs $b = 3, 4, 5$; R-squared very high (red)

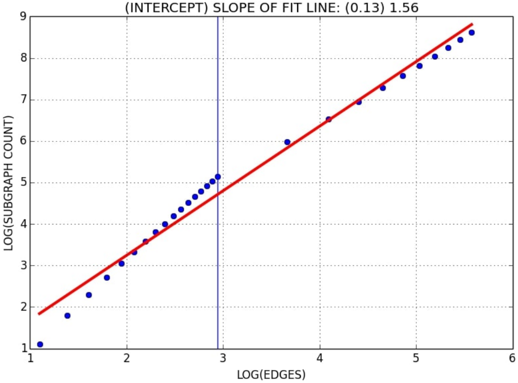
Scaling not robust to evolution in toy model (1)



- (Regime 1, $n < n^\star$) Grows as n-star
  e.g. $|M_3^{(3)}| \sim 2^{-1} m^2$, as $n^\star$ large

- (Regime 2, $n \geq n^\star$) Tends to n-complete
  e.g. $|M_3^{(3)}| \sim 2^{1/2} m^{3/2}$, as n large relative to $n^\star$

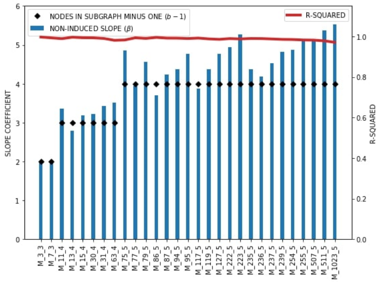
Scaling not robust to evolution in toy model (2)



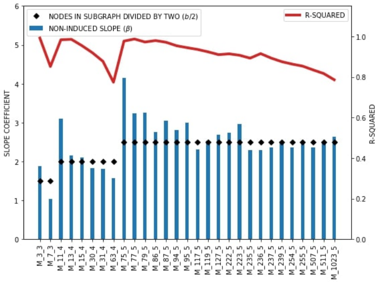
3-star

- Setting $n = 30$ and $n^{\star} = 20$, least squares on full-sample gives $\beta = 1.56$ and $R^2 = 0.983$ for 3-star subgraph counts

Evidence for power law $|M_a^{(b)}| = A\, m^{\beta}$, two regimes



Regime 1



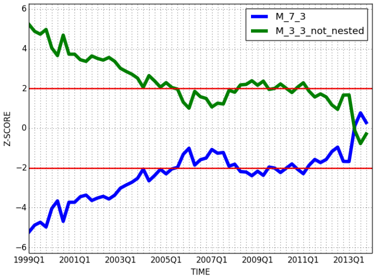
Regime 2

- In Regime 1, evidence that $\beta \approx b - 1$
- In Regime 2, evidence that $\beta \approx b/2$
- Which theoretical or real graphs have similar scaling? What could cause such a "phase transition"?

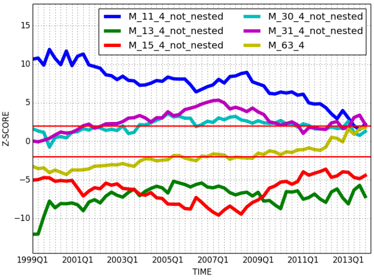
Which induced subgraphs are motifs?

- Do any subgraphs arise more (less) often than in G random? (Milo et al., 2002; many papers 2002–2004; Chen et al., 2013)
- Null ensemble: degree-preserving rewiring of edges to randomize observed G, control for observed lower-degree induced subgraphs by simulated annealing (Milo et al., 2002)
- Motifs unique local topologies, basic structural elements that *may* perform specialized functions individually or in interaction

The z-scores for 3-node and 4-node induced subgraphs



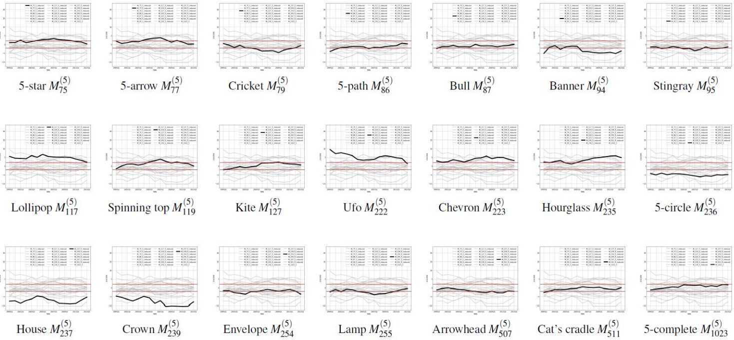
3-node



4-node

- The 3-star $\widetilde{M}_3^{(3)}$ less significant over time, triangle $M_7^{(3)}$ opposite interpretation, but density and clustering have fallen!
- The 4-star $\widetilde{M}_{11}^{(4)}$ is a strong motif for much of the sample, the 4-path $\widetilde{M}_{13}^{(4)}$ and tadpole $\widetilde{M}_{15}^{(4)}$ are strong anti-motifs

The z-scores for 5-node induced subgraphs



5-node

- Some heterogeneity across subgraphs, significance of motifs stable over time, curious given observed dynamics and scaling

Conclusions

- Frame real-world transportation network as a mathematical graph, investigate local topology, some features of random graphs but striking differences too (scaling, motifs)
- Quantitative evidence for power law between subgraph counts and edges, possible "phase transition" in underlying model
- Efficient detection of motifs on 3–5 nodes, some surprising quantitative results (density falls but "clustered" motifs more significant, time-varying scaling but motifs "stable" over time)

Extensions and speculation

- Develop a better qualitative and quantitative understanding of subgraph-based scaling in economic and transportation networks (better data, larger networks, larger subgraphs, bias-reducing econometrics, theoretical models, universality)

- Expand toolkit of subgraph-based "mesoscopic" measures, subgraphs as "fingerprints" of real-world networks? (network typology, components of econometric or statistical models)

- Do *small* subgraphs provide insight into the "reconstruction conjecture" (Kelly, 1957; Ulam, 1960) or analogous results? (are graphs uniquely determined by small/large subgraphs?)