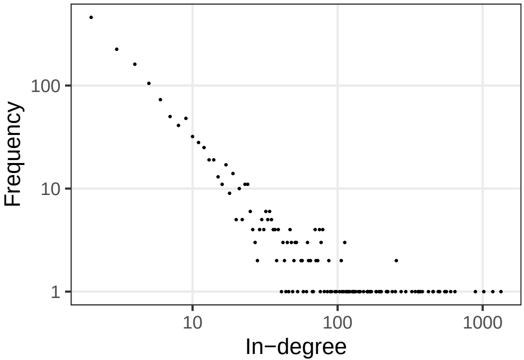
Evidencing preferential attachment in dependency network evolution

Clement Lee (Newcastle University, UK)

2025-06-26

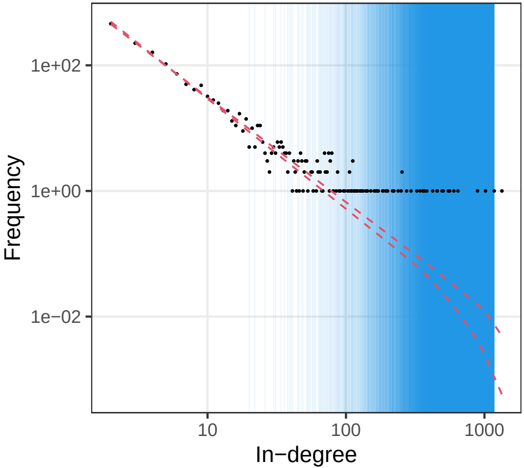
Introduction

- Interest in degree distribution
- Power law $\rightarrow$ Scale-freeness
- Associated with preferential attachment
  1. New node joins & brings new edges
  2. Node i selected with weight $g(d_i)$
  3. Barabási and Albert (1999): $g(d_i) = d_i$

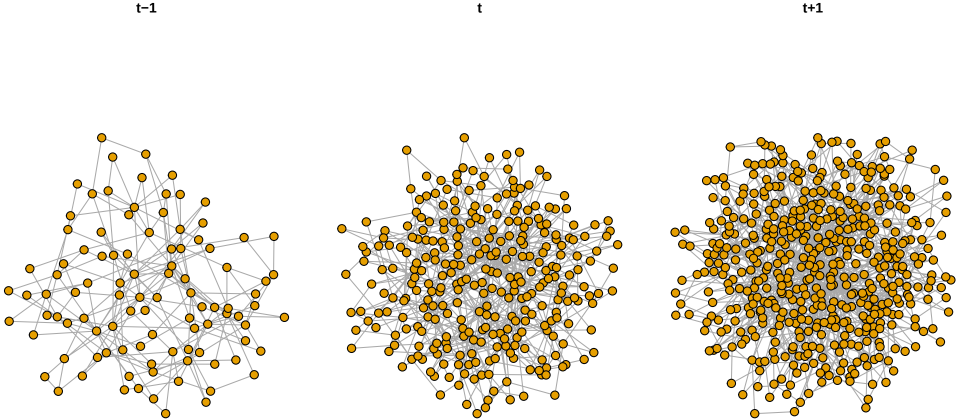


A few issues

1. General case $\nrightarrow$ power law
   - $g(d_i) = d_i^\alpha$ (Krapivsky and Redner 2001)
2. Other models $\rightarrow$ scale-freeness
   - Generalised random graphs (Hofstad 2016)
3. Debate on ubiquity
   - Broido and Clauset (2019): rare in real networks
   - Voitalov et al. (2019): rebuttal
   - Lee, Eastoe, and Farrell (2024): partial scale-freeness
4. Main cause: only snapshots available



Take a step back



Goals

If we have the evolution data:

1. Can we evidence the preferential attachment?
2. How do we model the data in a probabilistic / statistical way?
3. What is the precise form of the weight function?
    - $g(d_i) = d_i^\alpha$ or a more general form?

Daily increments of R packages on CRAN

```
##                from      to   add
## 1          glmmsr methods FALSE
## 2            GpGp     FNN  TRUE
## 3        pmxTools   stats FALSE
## 4  ceterisParibus   knitr  TRUE
## 5        pmxTools   xpose FALSE
## 6     mRchmadness   shiny FALSE
## 7          glmmsr    lme4 FALSE
## 8     mRchmadness   rvest FALSE
## 9           xpose   dplyr FALSE
## 10          xpose   utils FALSE
```

⋮

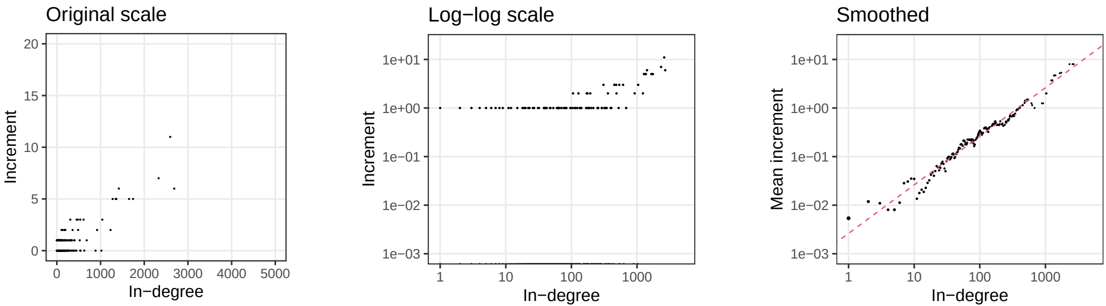
- From 2019-01-29 to 2024-12-31
- Obtained by `crandep::get_dep_all_packages()`
- ... which uses `tools::CRAN_package_db()`
- Change in `Imports` for one day
- Some new packages, some existing ones

Aggregating

```
##     previous date indegree increment count
## 1      2019-01-29        0         0  9315
## 2      2019-01-29        0         1     1
## 3      2019-01-29        1         0  1306
## 4      2019-01-29        2         0   459
## 5      2019-01-29        3         0   226
## 6      2019-01-29        4         0   161
## 7      2019-01-29        5         0   103
## 8      2019-01-29        6         0    75
## 9      2019-01-29        7         0    49
## 10     2019-01-29        8         0    40
```

⋮

Scatter plot

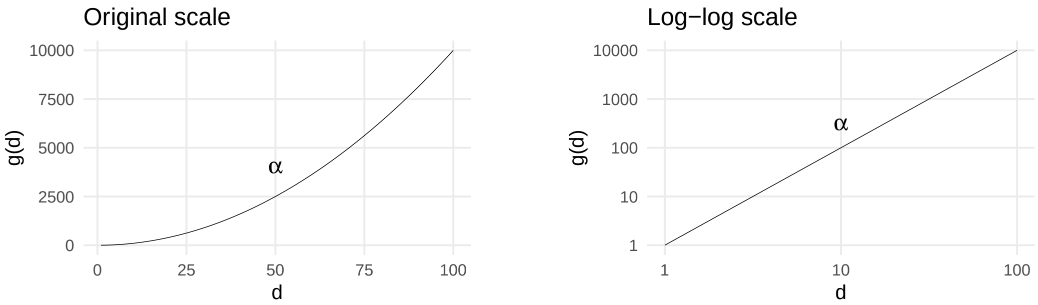


Regress increment on in-degree

- The **statistical** model based on preferential attachment
- Y_i: number of new edges / increments for node i
- $\{Y_i\} \quad \sim$ Multinomial with prob. $\left\{\frac{d_i^\alpha}{\sum_j d_j^\alpha}\right\}$, 1 new edge
- $\{Y_i\}|M \sim$ Multinomial with prob. $\left\{\frac{d_i^\alpha}{\sum_j d_j^\alpha}\right\}$, $M \sim \text{Poisson}(\mu)$ new edges
  - $\Rightarrow \{Y_i\} \sim$ Independent Poissons with mean $\left\{\frac{\mu \ d_i^\alpha}{\sum_j d_j^\alpha}\right\}$
  - $\log E(Y_i) = \alpha \log d_i + \underbrace{\log \mu - \log\left(\sum_j d_j^\alpha\right)}_{\text{constant}}$

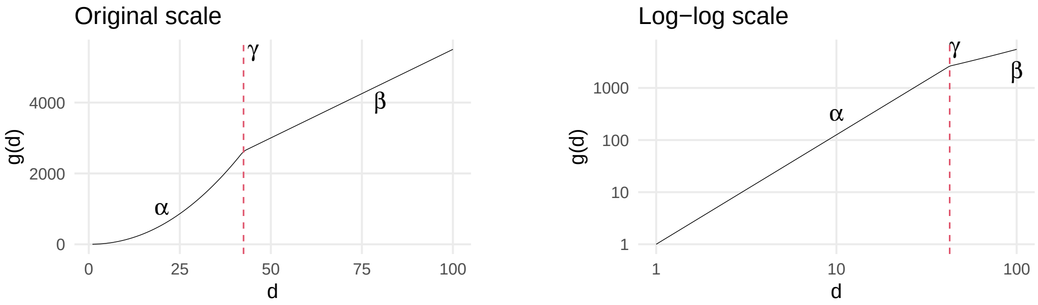
Revisiting weight function

- $g(d) = d^{\alpha}$
  - Inflexible tail behaviour of degree distribution (Oliveira and Spencer 2005; Rudas, Tóth, and Valkó 2007)
- Real data displays subtle tail behaviour
  - Heavy-tailed, but not as heavy as power law (Lee, Eastoe, and Farrell 2024)



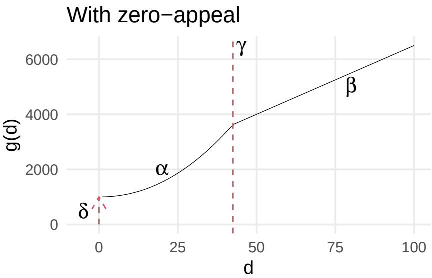
A more general one

- $g(d) = \begin{cases} d^{\alpha}, & d \leq \gamma, \\ \gamma^{\alpha} + \beta(d - \gamma), & d \geq \gamma \end{cases}$
  - Flexible heavy-tail behaviour (Boughen, Lee, and Palacios Ramirez 2025)

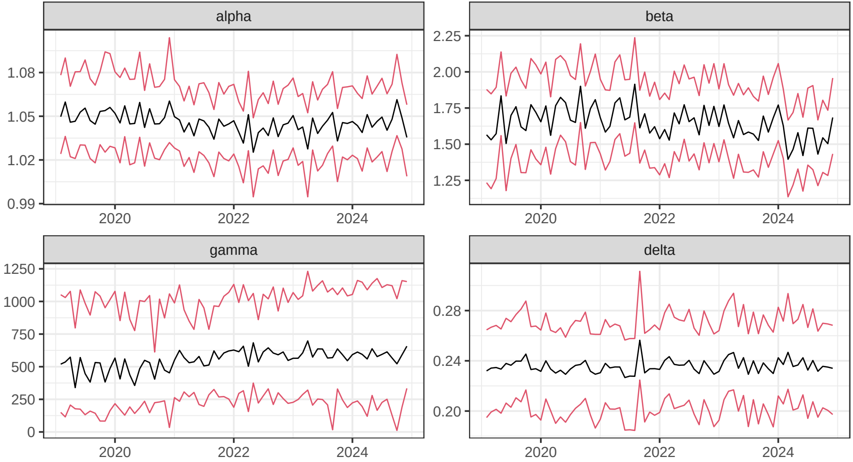


Modelling recipe

1. Regress increment on in-degree
   - $Y_i \sim \text{Poisson}\left(\frac{\mu\, g(d_i)}{\sum_j g(d_j)}\right)$
2. Choice of $g(d)$
   - Piecewise function (previous slide)
   - Power function for benchmark: $g(d) = d^{\alpha}$
   - Include parameter for zero-appeal δ
3. Bayesian inference for $(\alpha, \beta, \gamma, \delta, \mu)$
   - A set of parameters for each month of increments
   - Results of α most important

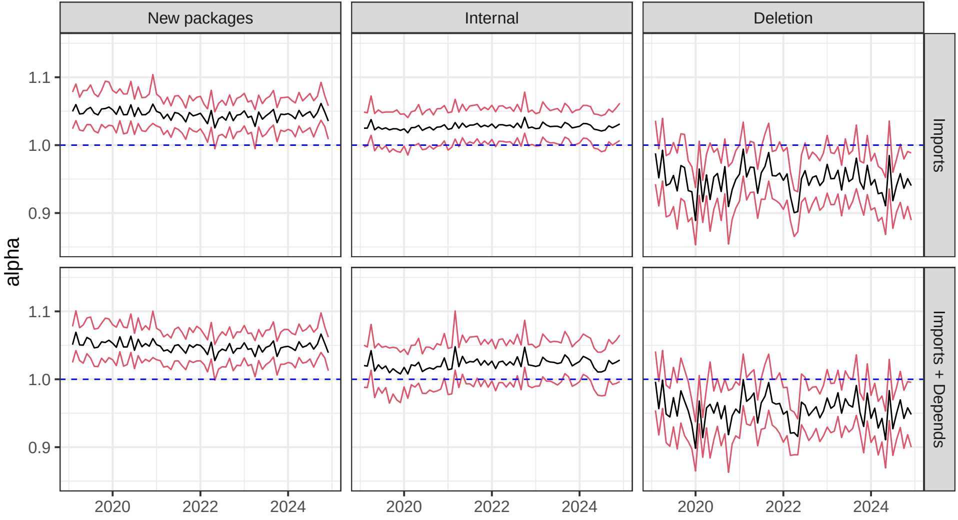


Piecewise function, `Imports`, new packages

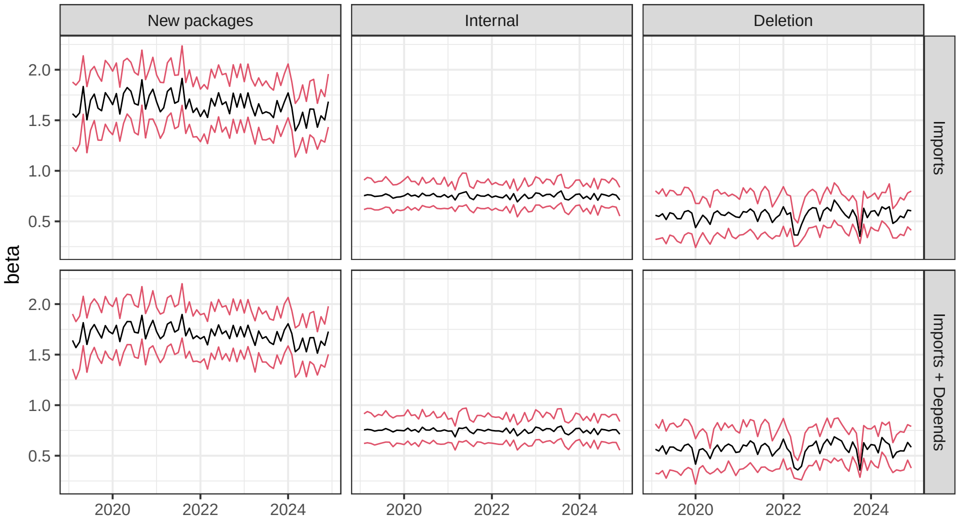


- Very similar results for α & δ using power function

α (super-/sub-linear?)

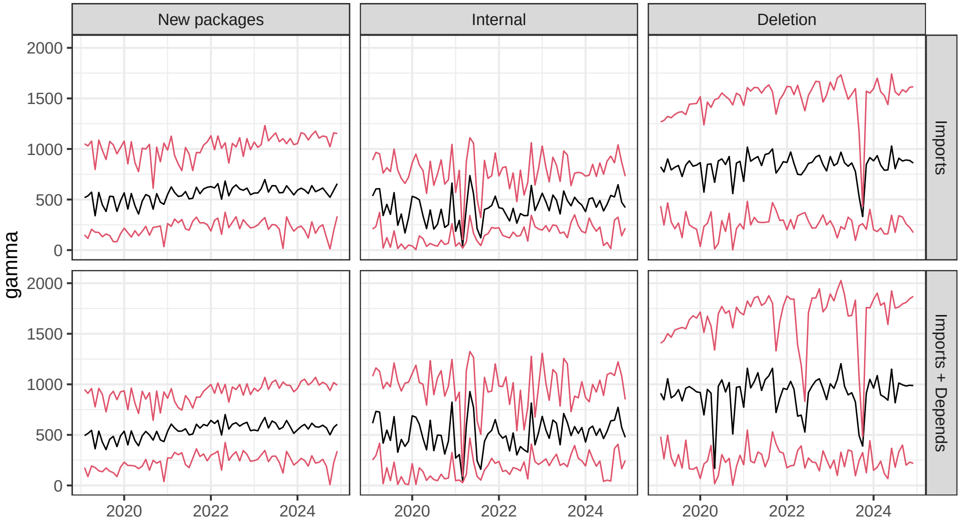


β



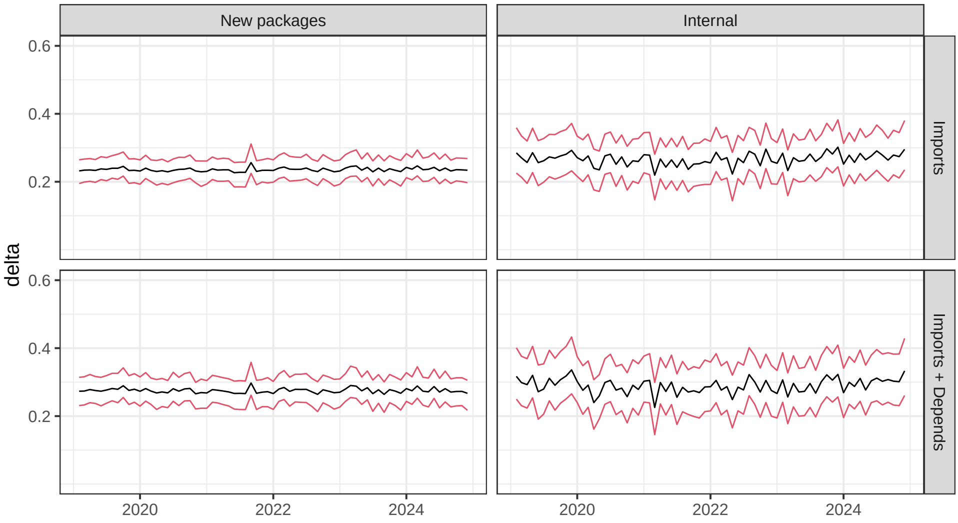
New packages
Internal
Deletion
Imports
Imports + Depends
beta
2.0
1.5
1.0
0.5
2020
2022
2024

γ



New packages
Internal
Deletion
Imports
Imports + Depends
gamma
2000
1500
1000
500
0
2020
2022
2024

δ


New packages
Internal
Imports
Imports + Depends
delta
0.6
0.4
0.2
2020
2022
2024

Summary

- Preferential attachment evidenced from increments of network evolution
  - increment vs in-degree on log-log scale
  - Poisson regression model
- Piecewise weight function of in-degree
  - allows partial super-/sub-linear preferential attachment
  - accommodates flexible heavy tail (of the degree distribution)
- Model applied to CRAN package dependencies
  - parameters stable over the observation period
  - adding edges: **Partial superlinear preferential attachment**
  - deleting edges: **Sublinear preferential detachment without zero-appeal**

Thank you

- This presentation's slides: https://bit.ly/sunbelt2025
- On partial scale-freeness: https://doi.org/10.1111/stan.12355
- On degree tail flexibility: https://doi.org/10.48550/arXiv.2506.18726

Bibliography

Barabási, Albert-László, and Réka Albert. 1999. "Emergence of Scaling in Random Networks." *Science* 286 (5439): 509–12. https://doi.org/10.1126/science.286.5439.509.

Boughen, Thomas, Clement Lee, and Vianey Palacios Ramirez. 2025. "Tail Flexibility in the Degrees of Preferential Attachment Networks." *ArXiv E-Prints*. https://doi.org/10.48550/arXiv.2506.18726.

Broido, A. D., and A. Clauset. 2019. "Scale-Free Networks Are Rare." *Nature Communications* 10 (1017). https://doi.org/10.1038/s41467-019-08746-5.

Hofstad, Remco van der. 2016. "Generalized Random Graphs." In *Random Graphs and Complex Networks*, 183–215. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. https://doi.org/10.1017/9781316779422.009.

Krapivsky, P. L., and S. Redner. 2001. "Organization of Growing Random Networks." *Physical Review E* 63 (6): 066123.

Lee, Clement, Emma F Eastoe, and Aiden Farrell. 2024. "Degree Distributions in