`backbone`: An R package for extracting the backbone of bipartite projections

Rachel Domagalski, Zachary Neal, Bruce Sagan
Michigan State University
NSF (#1851625 & #2016320)

PLoS ONE: https://doi.org/10.1371/journal.pone.0244363
All things backbone: www.zacharyneal.com/backbone

Backbone

The **backbone** of a bipartite projection is an unweighted or signed network that keeps only the **most important edges** of the projection.

The **`backbone`** R package provides different methods for extracting this backbone!

You can install it from CRAN and load it like this:

```
install.packages("backbone")
library(backbone)
```


@zpneal
@rdomagalski

Why use the `backbone` package?

Let's say we use a **co-sponsorship** network to measure political **collaboration** between senators.

When projected, **edge weights = number of co-sponsored bills**

BILLS

SENATORS

1	0	1

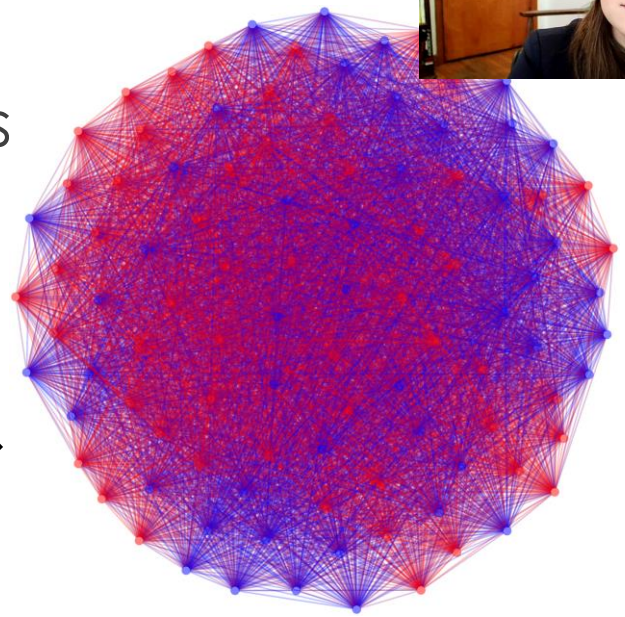
	100	

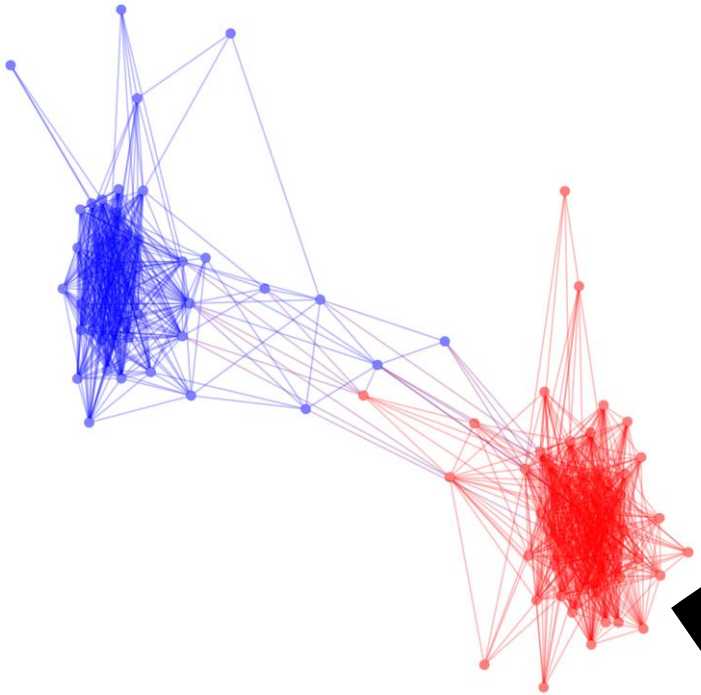
Edge weight depends on how many bills each person sponsors, and how many sponsors each bill tends to have.

@zpneal
@rdomagalski

Why use the `backbone` package?

Here's a visual of the co-sponsorship network of US Senators in the 114th session. An edge was kept if two Senators co-sponsored at least one bill together.





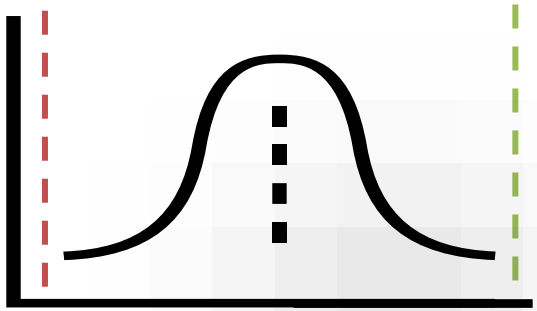
Here's that same network using the `backbone` package's SDSM model! Most important edges are found to exist within party.

@zpneal
@rdomagalski

How Does It Work?

Backbone uses **statistical tests** to compare an edge's observed weight in the bipartite projection to the **distribution of its weights** expected under a **null model**.

An edge's observed weight is **statistically significant** if it is in the upper or lower tail of the distribution.



Upper tail = **positive** edge in the backbone
Lower tail = **negative** edge in the backbone

@zpneal
@rdomagalski

Three Different Types of Null Models

Compares the observed edge weight to a distribution of random bipartite graphs where…

Hypergeometric: `hyperg`

…the row sums exactly match the observed bipartite, *but* column sums are allowed to vary.

Stochastic Degree Sequence Model: `sdsm`

…the row sums *and* column sums are allowed to vary, but match on average.

Fixed Degree Sequence Model: `fdsm`

…the row sums and column sums *exactly* match the observed bipartite.

@zpneal
@rdomagalski

Using the backbone package

Your Data

- Adjacency matrix
- A sparse `Matrix` object
- Bipartite `igraph` object
- Bipartite `network` object
- An edgelist

Apply Null Model

`hyperg`

`sdsm`

`fdsm`

Get `backbone` Object

`positive` - matrix with entries equal to the *probability* of an edge under the null distribution having weight equal to or *above* the weight observed in the projection.
`negative` - *probability* of equal to or *below*…

Extract Backbone

Plug in `backbone` object to **`backbone.extract()`** to find a positive or signed backbone at a given significance level.

@zpneal
@rdomagalski

Extracting the Backbone

```
backbone.extract(backbone,
                 signed = TRUE,
                 alpha = .01,
                 fwer = "none",
                 class = "original",
                 narrative = TRUE)
```

backbone: a backbone class object as returned by `hyperg`, `sdsm`, `fdsm`

signed: TRUE for a signed backbone, FALSE for a positive backbone

alpha: significance test value (always a two-tailed test)

fwer: Family-wise error rate test correction. Choose from "none", "Bonferroni", and "Holm"

class: type of graph object you'd like back: matrix, igraph, network, edgelist

narrative: suggested manuscript text and citations

@zpneal
@rdomagalski

Hypergeometric Distribution

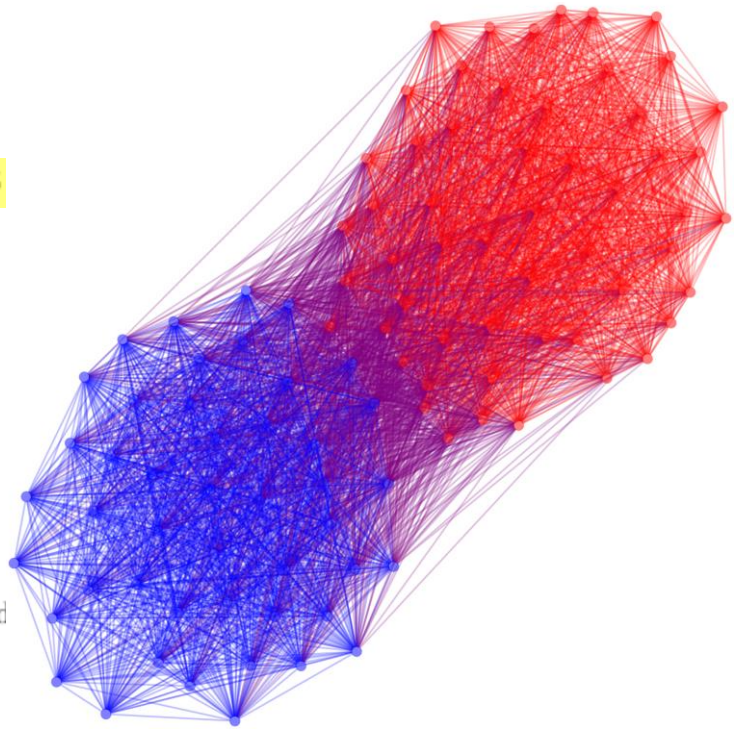
Senate co-sponsorship: fixes the number of bills that each senator sponsored, while allowing each bill to be sponsored by a varying number of senators.

On the Senate co-sponsorship data from the 114th session, we get:

```
hyperg_probs <- hyperg(data)
hyperg_bb <- backbone.extract(hyperg_probs
                              alpha = .01)
```

"backbone" class object that contains the probabilities

Plug entire thing into the `backbone.extract` function



Tumminello M, Miccichè S, Lillo F, Piilo J, Mantegna RN. Statistically Validated Networks in Bipartite Complex Systems. PLOS ONE. 2011;6(3):e17994. doi:10.1371/journal.pone.0017994.

Neal Z. Identifying statistically significant edges in one-mode projections. Social Network Analysis and Mining. 2013;3(4):915–924. doi:10.1007/s13278-013-0107-y.

@zpneal
@rdomagalski

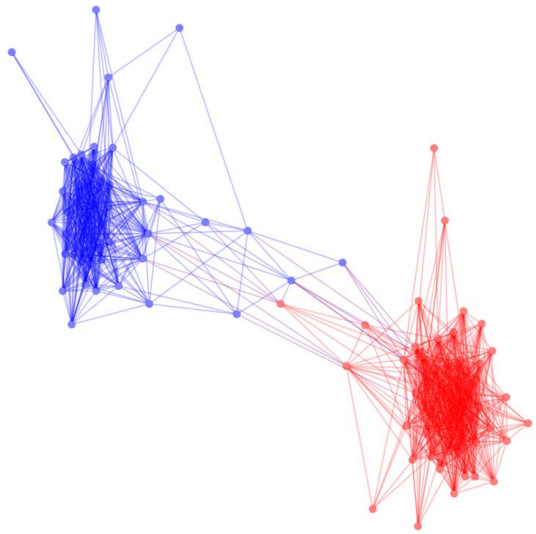
Stochastic Degree Sequence Model

Senate co-sponsorship: compares the observed values to the distribution where each senator sponsors roughly the same number of bills, and each bill is sponsored by roughly the same number of people.

```
sdsm <- sdsm(data)
sdsm_bb <- backbone.extract(sdsm, alpha = .01, narrative = TRUE)
```

1) computes probabilities for each cell in the bipartite matrix to be equal to 1

2) applies the **Poisson Binomial distribution** to the weighted projection.



@zpneal
@rdomagalski

Stochastic Degree Sequence Model

Senate co-sponsorship: compares the observed values to the distribution where each senator sponsors roughly the same number of bills, and each bill is sponsored by roughly the same number of people.

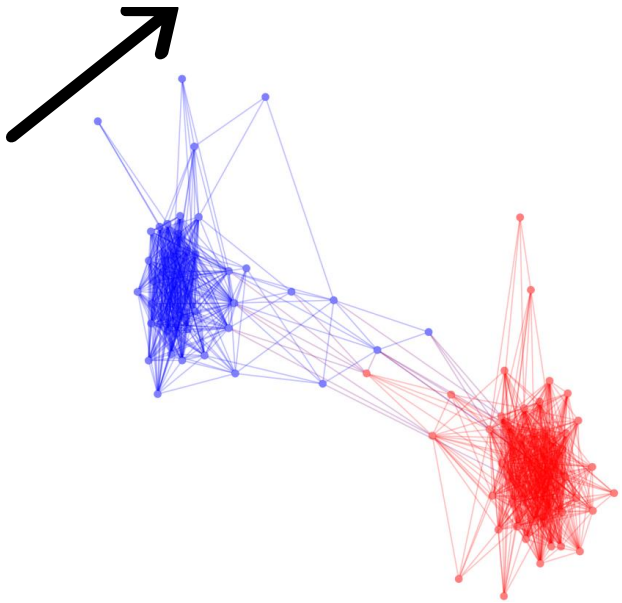
```
sdsm <- sdsm(data)
sdsm_bb <- backbone.extract(sdsm, alpha = .01, narrative = TRUE)
```

```
Suggested manuscript text and citations:

From a bipartite graph containing 100 agents and 3589 arti
facts, we obtained the weighted bipartite projection, then
 extracted its signed backbone using the backbone package
 (Domagalski, Neal, & Sagan, 2021). Edges were retained in
 the backbone if their weights were statistically signific
ant (alpha = 0.01) by comparison to a null Stochastic Degr
ee Sequence Model (Neal, 2014).

Domagalski, R., Neal, Z. P., and Sagan, B. (2021). backbon
e: An R Package for Backbone Extraction of Weighted Graph
s. PLoS ONE. https://doi.org/10.1371/journal.pone.0244363

Neal, Z. P. (2014). The backbone of bipartite projections:
 Inferring relationships from co-authorship, co-sponsorshi
p, co-attendance and other co-behaviors. Social Networks,
 39, 84-97. https://doi.org/10.1016/j.socnet.2014.06.001
```



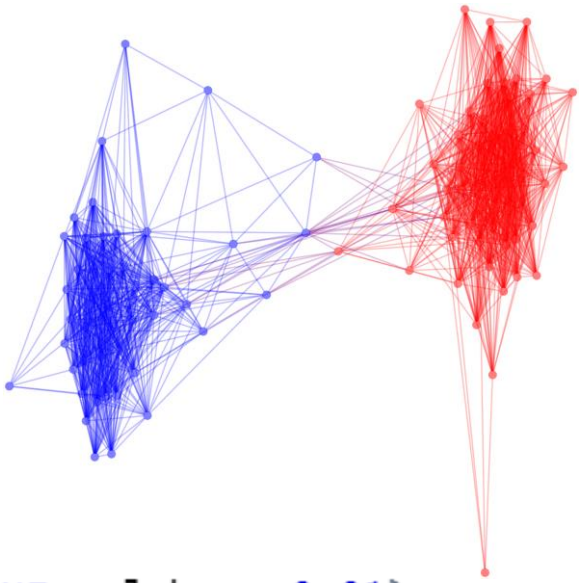
@zpneal
@rdomagalski

Fixed Degree Sequence Model

Senate co-sponsorship: compares the observed values to a distribution where each senator sponsors the exact same number of bills and each bill is sponsored by the exact same number of people.

Samples this distribution by applying the **curveball** algorithm to the bipartite graph numerous times.

Choose the number of times it's applied with the "**`trials`**" parameter (more trials, more accurate)

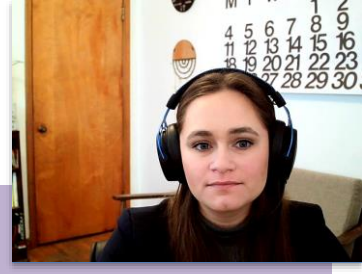


```
fdsm <- fdsm(data, trials = 1000)
fdsm_bb <- backbone.extract(fdsm, signed = TRUE, alpha = 0.01)
```

Zweig KA, Kaufmann M. A systematic approach to the one-mode projection of bipartite graphs. Social Network Analysis and Mining. 2011;1(3):187–218. doi:10.1007/s13278-011-0021-0.

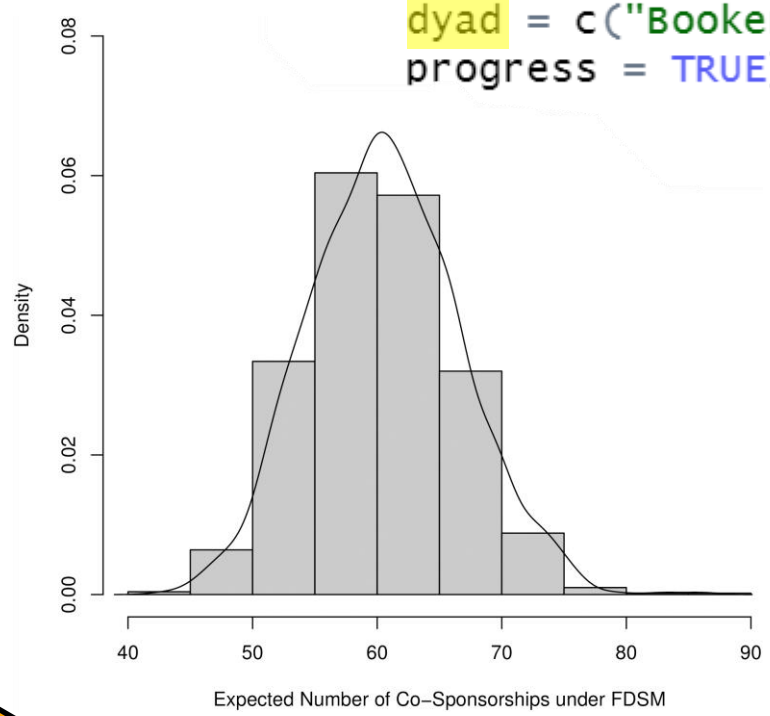
@zpneal
@rdomagalski

Fixed Degree Sequence Model



Senate co-sponsorship: compares the observed values to a distribution where each senator sponsors the exact same number of bills and each bill is sponsored by the exact same number of people.

```
fdsm <- fdsm(data, trials = 1000,
             dyad = c("Booker, C. (NJ-D)", "Warren, E. (MA-D)"),
             progress = TRUE)
```



Can also get the distribution for a chosen **dyad.**

```
fdsm$dyad_values
```

We can visualize this data in a histogram.

← Their actual number of cosponsorships is 98.

@zpneal
@rdomagalski

Coming soon!

Early February 2021: backbone 1.3.0

- Easier model selection / guidance
- Faster SDSM computations
- Bells & Whistles

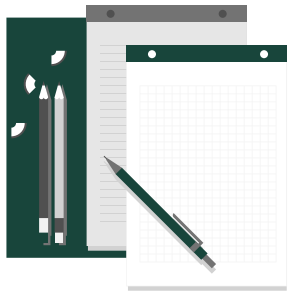
Summer 2021: backbone 2.0.0

- Models for general weighted graphs

(not necessarily bipartite projections)

Manuscript to aid in model selection

- Accuracy of probabilities generated in SDSM
- Effects of degree distributions and significance levels
- Ability to recover group structure



@zpneal
@rdomagalski

Backbone in PLoS ONE

PLOS ONE

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Backbone: An R package for extracting the backbone of bipartite projections

Rachel Domagalski, Zachary P. Neal, Bruce Sagan

Published: January 6, 2021 • https://doi.org/10.1371/journal.pone.0244363

Article | Authors | Metrics | Comments | Media Coverage | Peer Review

Abstract

Introduction

Backbone extraction methods

Applying backbone to the US Senate

Discussion

Supporting information

References

Reader Comments (0)

Media Coverage (0)

Figures

Abstract

Bipartite projections are used in a wide range of network contexts including politics (bill co-sponsorship), genetics (gene co-expression), economics (executive board co-membership), and innovation (patent co-authorship). However, because bipartite projections are always weighted graphs, which are inherently challenging to analyze and visualize, it is often useful to examine the 'backbone,' an unweighted subgraph containing only the most significant edges. In this paper, we introduce the `R` package `backbone` for extracting the backbone of weighted bipartite projections, and use bill sponsorship data from the 114th session of the United States Senate to demonstrate its functionality.

Figures

Citation: Domagalski R, Neal ZP, Sagan B (2021) Backbone: An R package for extracting the backbone of bipartite projections. PLoS ONE 16(1): e0244363. https://doi.org/10.1371/journal.pone.0244363

Editor: Celine Rozenblat, University of Lausanne, SWITZERLAND

Received: July 29, 2020; **Accepted:** December 8, 2020; **Published:** January 6, 2021

Copyright: © 2021 Domagalski et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The data and code necessary to replicate the examples in this paper are available at https://osf.io/myje5/.

Funding: ZN and BS received funding from the National Science Foundation (#1851625 & #2016320). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Learn in detail about each null model and the mathematics behind them

Data and code available to replicate all the results and examples

https://doi.org/10.1371/journal.pone.0244363

www.zacharyneal.com/backbone

github.com/domagal9/backbone

Tweet us any questions or comments!

THANK YOU!

@zpneal

@rdomagalski