

# THE YERSINIA PESTIS METABOLIC NETWORK VERTEX-DEGREE DISTRIBUTION APPROXIMATES A POWER-LAW

Author: J. Horner, jhorner@cybermesa.com

Last modified: 11 October 2022/1210 CT

Platform: *Mathematica* (Wolfram 2022)

## Abstract

*Yersinia pestis* is the bacterium that causes the bubonic plague. Here I show that the distribution of vertex-degree in the *Yersinia pestis* metabolic network approximates a power-law. A network with a power-law vertex-degree distribution is robust under random-vertex deletion. This suggests that there are relatively few efficacious therapeutic targets in the metabolic network of *Y. pestis*.

## 1.0 Introduction

### 1.1 Some definitions

A *metabolic network* is the complete set of metabolic and physical processes that determine the physiological and biochemical properties of a cell. As such, these networks comprise the chemical reactions of metabolism, the metabolic pathways, as well as the regulatory interactions that guide these reactions (Nelson and Cox 2005, esp. Chaps. 14-18).

*Yersinia pestis* is the bacterium that causes the bubonic plague (Ryan and Ray 2004, pp. 484–488).

*Power-law distribution.* If a variable  $P$  is related to a variable  $Q$  (in this case  $\log$  vertex-degree) by Eq. 1,

$$\log P = \alpha - \tau \log Q \quad \text{Eq. 1}$$

where  $\alpha$  is a real constant and  $\tau$  is a positive real, we say that  $P$  is a *power-law function of Q* ([Newman 2005, p. 331] or that  $P$  is a *power-law distribution function of Q*. The converse also holds.

The previous paragraph implies that  $P$  is a power-law function of  $Q$  if and only if  $\log P$  is a linear function of  $\log Q$ . The question of how closely a given distribution approximates a power-law can therefore be answered by evaluating how well a given distribution approximates a linear relation between  $\log P$  and  $\log Q$ . This is the approach taken in Section 2.0, below.

In power-law-distributions  $\tau$  must have a value greater than 2 for the mean of that distribution to be defined (Newman 2005, p. 332);  $\tau$  must have a value greater than 3 for the variance to be defined

(Newman 2005, p. 332). If  $\tau$  is 2 or less (as is the case in the calculation below), therefore, neither the mean nor the variance of the distribution is defined.

Distributions in which neither the mean nor the variance are defined do not support conventional statistical inference methods (Kanji 2006) that require at least the mean or the variance of the distribution of interest to be defined.

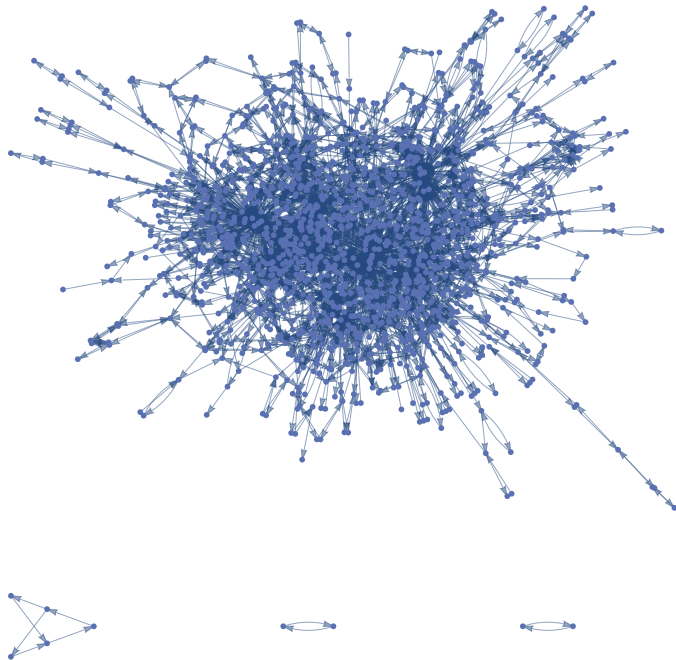
## 1.2 Networks with power-law vertex-degree distributions

Most vertices in a network whose vertex-degree distribution is a power-law are connected to very few other vertices in that network. A network with a power-law vertex-degree distribution is robust under random-vertex deletion (Barabási 2016, esp. Chap. 8).

## 2.0 The calculation

We first obtain a sample of the *Y. pestis* metabolic network from the *Mathematica* archives, then produce a graphic representation of of that network.

```
In[43]:= g = ExampleData[{"NetworkGraph", "MetabolicNetworkYersiniaPestis"}]
Out[43]=
```



Get some descriptive information on *g* from the *Mathematica* archives.

```
In[44]:= ExampleData[{"NetworkGraph", "MetabolicNetworkYersiniaPestis"}, "LongDescription"] //
TraditionalForm
Out[44]//TraditionalForm=
Metabolic cellular network data for Yersinia Pestis
```

Next, we extract a list of vertex degrees from the sample.

```
In[45]:= degreeList = VertexDegree[g]
Out[45]=
{6, 3, 4, 4, 4, 3, 2, 3, 1, 4, 4, 4, 4, 7, 2, 3, 29, 3, 3, 4, 3, 2, 5, 4, 4, 4, 2, 4, 4, 4, 3, 2, 2,
1, 5, 7, 4, 4, 4, 2, 4, 4, 4, 4, 4, 3, 4, 4, 4, 2, 4, 4, 2, 4, 4, 2, 4, 6, 1, 3, 3, 4, 2, 6, 2,
3, 1, 3, 6, 3, 4, 4, 6, 4, 4, 3, 4, 4, 3, 4, 3, 4, 4, 4, 3, 4, 5, 3, 4, 2, 4, 3, 4, 4, 2, 4, 8,
3, 4, 4, 2, 4, 5, 3, 6, 2, 4, 4, 4, 3, 4, 4, 5, 4, 4, 3, 4, 2, 4, 4, 4, 2, 4, 4, 4, 2, 3, 5, 3,
3, 4, 2, 4, 3, 6, 3, 3, 4, 6, 3, 2, 4, 2, 4, 2, 4, 2, 4, 7, 4, 6, 4, 4, 4, 6, 6, 4, 4, 4, 4, 4,
4, 3, 4, 9, 4, 4, 4, 2, 3, 4, 2, 4, 2, 3, 2, 5, 3, 4, 4, 4, 3, 3, 2, 4, 2, 4, 2, 4, 10, 2, 4,
4, 4, 4, 4, 3, 2, 3, 5, 3, 2, 2, 12, 4, 3, 3, 3, 3, 3, 6, 4, 4, 3, 1, 4, 4, 4, 4, 2, 4, 2, 6,
2, 4, 2, 4, 2, 6, 2, 6, 3, 4, 4, 4, 4, 6, 4, 4, 5, 36, 4, 5, 6, 4, 4, 4, 5, 4, 4, 5, 4, 4, 4,
5, 4, 4, 4, 4, 4, 2, 3, 10, 4, 3, 3, 4, 3, 4, 3, 2, 3, 3, 2, 3, 4, 5, 4, 2, 3, 3, 3, 3, 4, 4,
5, 3, 4, 5, 4, 3, 4, 3, 3, 6, 3, 3, 4, 6, 3, 4, 3, 4, 4, 4, 2, 4, 2, 3, 5, 4, 5, 4, 6, 4, 4,
4, 3, 4, 4, 3, 3, 5, 4, 3, 3, 4, 4, 4, 2, 4, 4, 5, 4, 5, 3, 2, 5, 3, 4, 3, 2, 5, 2, 2, 3, 4,
4, 4, 4, 5, 4, 4, 4, 4, 8, 4, 3, 2, 8, 2, 5, 7, 4, 4, 5, 14, 4, 5, 4, 4, 5, 4, 4, 3, 4, 3, 2,
2, 4, 2, 3, 2, 2, 4, 3, 2, 4, 2, 5, 16, 5, 4, 4, 4, 4, 4, 6, 4, 4, 3, 6, 2, 4, 2, 2, 6, 4, 4,
4, 4, 4, 4, 5, 3, 4, 8, 3, 5, 3, 3, 3, 4, 2, 4, 5, 4, 6, 4, 4, 5, 3, 138, 4, 4, 6, 4, 5, 4, 4,
3, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 6, 4, 4, 5, 4, 6, 4, 4, 6, 4, 5, 4, 4, 4, 6, 4,
4, 6, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 43, 6, 4, 3, 7, 6, 6, 4, 4, 4, 181, 4, 6, 4, 6, 5, 4, 4,
4, 4, 7, 4, 6, 8, 6, 4, 3, 4, 4, 6, 4, 8, 6, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4,
4, 4, 4, 4, 5, 4, 6, 6, 6, 4, 6, 4, 4, 6, 4, 4, 8, 5, 4, 4, 4, 6, 4, 4, 4, 5, 4, 4, 6, 4, 4,
6, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 4, 7, 7, 7, 7, 7, 7, 4, 6, 4, 7, 2, 4, 7, 5, 6, 5, 4, 4, 4,
9, 4, 4, 4, 55, 4, 4, 4, 5, 3, 5, 5, 4, 12, 5, 4, 6, 1, 4, 39, 4, 4, 5, 4, 4, 4, 4, 4, 5, 3,
3, 4, 3, 4, 4, 2, 6, 4, 4, 3, 2, 4, 2, 4, 8, 5, 4, 4, 4, 2, 4, 17, 2, 4, 4, 4, 2, 2, 2, 2, 5,
3, 4, 4, 2, 4, 8, 4, 2, 4, 4, 8, 4, 4, 4, 2, 5, 7, 4, 4, 2, 4, 2, 4, 4, 4, 2, 4, 4, 2, 4, 4,
2, 2, 4, 9, 4, 4, 2, 4, 4, 10, 2, 2, 2, 4, 4, 4, 2, 2, 4, 8, 2, 2, 4, 4, 3, 9, 4, 4, 5, 4, 7,
4, 4, 7, 3, 4, 3, 4, 3, 5, 3, 10, 6, 4, 3, 2, 4, 4, 2, 4, 4, 4, 2, 4, 4, 6, 3, 2, 2, 2, 4, 4,
3, 9, 5, 9, 4, 4, 6, 4, 29, 3, 3, 4, 2, 4, 4, 4, 4, 4, 254, 4, 6, 4, 4, 3, 5, 3, 3, 3, 4, 4,
4, 3, 4, 3, 5, 4, 4, 4, 4, 4, 4, 4, 3, 4, 3, 5, 5, 3, 5, 4, 5, 5, 4, 4, 4, 4, 4, 3, 4,
4, 5, 4, 4, 4, 4, 4, 4, 3, 4, 4, 3, 4, 5, 4, 5, 4, 3, 4, 4, 4, 4, 5, 5, 5, 5, 4, 4, 4, 3,
4, 3, 4, 3, 4, 4, 5, 6, 3, 6, 5, 5, 5, 5, 5, 4, 4, 4, 6, 4, 3, 9, 5, 1, 4, 3, 2, 8, 4, 2,
2, 4, 2, 2, 6, 4, 4, 3, 4, 3, 5, 24, 3, 3, 4, 4, 4, 7, 5, 4, 4, 4, 4, 6, 1, 85, 4, 4, 4,
4, 4, 4, 3, 5, 4, 2, 4, 4, 4, 3, 4, 4, 4, 2, 4, 5, 6, 3, 5, 24, 5, 2, 2, 6, 2, 6, 8, 3, 2, 5,
4, 4, 4, 6, 4, 4, 2, 2, 2, 2, 7, 3, 4, 4, 1, 10, 4, 4, 3, 3, 4, 2, 7, 4, 4, 4, 4, 2, 4, 3, 5,
2, 2, 11, 4, 4, 2, 4, 4, 3, 2, 4, 2, 4, 4, 2, 2, 7, 2, 2, 4, 3, 3, 2, 4, 4, 2, 2, 2, 1, 61,
5, 5, 5, 4, 59, 4, 5, 4, 5, 5, 6, 4, 58, 5, 4, 5, 4, 58, 4, 4, 5, 5, 5, 4, 4, 56, 4, 4, 4, 3,
4, 4, 3, 4, 12, 4, 16, 4, 4, 4, 4, 4, 1, 2, 3, 2, 2, 9, 5, 5, 3, 1, 4, 2, 2, 4, 4, 2, 2, 2,
8, 2, 6, 11, 4, 5, 4, 4, 3, 3, 4, 4, 4, 4, 4, 3, 3, 4, 3, 4, 2, 3, 2, 1, 4, 4, 4, 4, 2, 1, 4,
4, 3, 8, 3, 2, 4, 4, 1, 3, 2, 3, 2, 4, 2, 4, 4, 17, 4, 4, 3, 4, 4, 4, 4, 4, 4, 1, 1, 4, 2,
4, 3, 2, 3, 3, 3, 1, 2, 5, 1, 7, 4, 4, 7, 4, 11, 4, 23, 4, 4, 4, 4, 4, 2, 11, 5, 4, 2, 5,
2, 2, 2, 2, 2, 2, 4, 4, 4, 2, 1, 2, 3, 2, 2, 1, 5, 2, 4, 9, 3, 3, 1, 2, 1, 7, 1, 1, 9, 2, 2,
3, 5, 3, 3, 3, 2, 2, 1, 1, 5, 2, 2, 6, 2, 4, 4, 2, 3, 4, 4, 2, 4, 4, 3, 5, 4, 4, 6, 5, 3, 2,
4, 4, 1, 3, 4, 4, 4, 4, 3, 1, 2, 4, 4, 4, 4, 4, 3, 3, 2, 4, 3, 3, 1, 3, 2, 2, 1, 1, 3, 2,
3, 3, 3, 3, 3, 2, 4, 2, 14, 2, 3, 7, 4, 2, 2, 5, 3, 9, 2, 5, 3, 2, 7, 2, 4, 2, 2, 2, 4, 2, 3,
3, 2, 4, 5, 3, 3, 5, 5, 4, 3, 2, 2, 2, 2, 3, 7, 2, 7, 4, 4, 3, 2, 3, 2, 6, 3, 3, 2, 4, 11, 2,
4, 4, 4, 2, 2, 2, 2, 1, 2, 2, 3, 6, 4, 3, 4, 3, 108, 4, 4, 4, 5, 4, 4, 9, 9, 2, 2, 2, 4, 2,
```

```
3, 4, 12, 4, 3, 3, 4, 1, 2, 4, 2, 2, 2, 2, 1, 4, 4, 2, 6, 2, 2, 2, 2, 3, 2, 2, 2, 3, 3, 2, 6,
4, 2, 2, 2, 2, 1, 68, 4, 2, 4, 35, 3, 2, 2, 5, 2, 7, 2, 4, 3, 14, 5, 8, 4, 4, 7, 6, 5, 4, 2,
2, 6, 2, 4, 2, 2, 4, 2, 4, 1, 9, 1, 3, 4, 4, 3, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 14, 1, 1, 1, 2, 1,
1, 1, 1, 1, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 4, 1, 1}
```

```
In[46]:= Length[degreelist]
```

```
Out[46]=
```

```
1453
```

Bin the magnitudes using a 0.25-magnitude-unit bin width.

```
In[47]:= binlist = Transpose[{Range[5, 10, 0.25], BinCounts[degreelist, {4.75, 10, 0.25}]}]
```

```
Out[47]=
```

```
{ {5., 0}, {5.25, 116}, {5.5, 0}, {5.75, 0}, {6., 0}, {6.25, 74}, {6.5, 0},
  {6.75, 0}, {7., 0}, {7.25, 30}, {7.5, 0}, {7.75, 0}, {8., 0}, {8.25, 16},
  {8.5, 0}, {8.75, 0}, {9., 0}, {9.25, 14}, {9.5, 0}, {9.75, 0}, {10., 0} }
```

binlist is a list, denoted by the outermost curly brackets shown above, of a set of ordered pairs of {degree-magnitude, binned-count} values. Note that the counts in some elements of binlist are zero.

Let's translate the counts in binlist to the logarithms (base 10) of those counts, in the process excluding any bins in binlist that have a count of zero (the logarithm of zero is undefined), to obtain a "scrubbed" list, scrublist.

To do this, we first initialize scrublist to the null list:

```
In[48]:= scrublist = {};
```

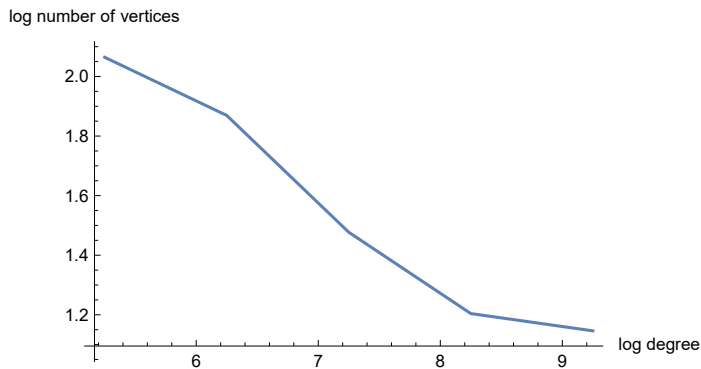
Next, we append to scrublist all (degree-magnitude, log binned-count) pairs that have nonzero bin-counts:

```
In[49]:= For[i = 1, i < Length[binlist] + 1, i++,
  If[binlist[[i]][[2]] != 0, AppendTo[scrublist, {binlist[[i]][[1], Log10[binlist[[i]][[2]]}]]]]
```

Now plot scrublist.

```
In[50]:= ListLinePlot[scrublist, AxesLabel → {"log degree", "log number of vertices"}]
```

```
Out[50]=
```



This plot suggests that there is approximately a linear relationship between log vertex-degree and log number of vertices with a given degree.

Let's calculate what the data in scrublist tell us about the linear form shown in Eq. 1. To do this, we first compute a function, lm, to fit scrublist, by linear regression:

```
In[51]:= lm = LinearModelFit[scrublist, {x}, x]
```

```
Out[51]=
```

```
FittedModel [ 3.366 - 0.250177 x ]
```

Note that the coefficient (0.250177) of  $x$  in lm is in log-log space. To convert that value to the corresponding value of  $\tau$  in Eq. 1, we must compute the antilog of 0.250177, which is 1.779. Because that value is less than 2, neither the mean nor variance of the vertex-degree distribution is defined.

To determine how well lm fits the data in scrublist, let's compute R-squared, and the significance probabilities for the parameter t-statistics, for the fit of lm to scrublist:

```
In[52]:= lm["RSquared"]
```

```
Out[52]=
```

```
0.956123
```

```
In[53]:= lm["ParameterPValues"]
```

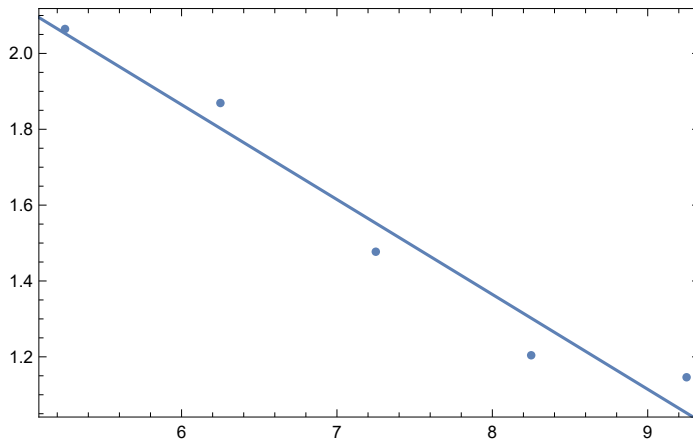
```
Out[53]=
```

```
{0.000679131, 0.00395329}
```

Show the data together with the fitted function.

```
In[54]:= Show[ListPlot[scrublist], Plot[lm[x], {x, 0, 10}], Frame -> True]
```

```
Out[54]=
```



Compute the residuals of the fit.

```
In[55]:= lm["FitResiduals"]
```

```
Out[55]=
```

```
{0.0118919, 0.0668428, -0.0750905, -0.0979146, 0.0942706}
```

Generate an analysis of variance table

```
In[56]:= lm["ANOVATable"]
```

```
Out[56]=
```

	DF	SS	MS	F-Statistic	P-Value
x	1	0.625886	0.625886	65.3731	0.00395329
Error	3	0.0287222	0.00957406		
Total	4	0.654608			

Extract the design matrix and residuals.

```
In[57]:= {desmat, resids} = lm[{"DesignMatrix", "FitResiduals"}];
```

Compute White's heteroscedasticity-consistent covariance estimate (White 1980):

```
In[58]:= (wc = With[{inv = Inverse[Transpose[desmat].desmat], xresid = resids * desmat},  
  inv.Transpose[xresid].xresid.inv]) // MatrixForm
```

```
Out[58]//MatrixForm=
```

$$\begin{pmatrix} 0.0209618 & -0.00318502 \\ -0.00318502 & 0.000501687 \end{pmatrix}$$

Compare with the covariance assuming homoscedasticity:

```
In[59]:= lm["CovarianceMatrix"] // MatrixForm
```

```
Out[59]//MatrixForm=
```

$$\begin{pmatrix} 0.0522385 & -0.00694119 \\ -0.00694119 & 0.000957406 \end{pmatrix}$$

Compare standard errors based on the two covariance estimates:

```
In[60]:= Sqrt [Diagonal [wc ]
Out[60]=
{0.144782, 0.0223984}

In[61]:= Im ["ParameterErrors"]
Out[61]=
{0.228557, 0.030942}
```

### 3.0 Conclusions

The analyses in Section 2.0 are strong evidence that the distribution of vertex-degree in the metabolic network of *Y. pestis* approximates a power-law function. When this distribution is cast in the form of Eq. 1, the value of  $\tau$  is 1.779, which implies neither the mean nor variance of the vertex-degree distribution is defined.

In addition, a network with a power-law vertex-degree distribution is robust under random-vertex deletion. This suggests that there are relatively few efficacious therapeutic targets in the metabolic network of *Y. pestis*.

### 4.0 References

Barabasi A-L. (2016). *Network Science*. Cambridge.

Kanji GK. (2006). *100 Statistical Tests*. Third Edition. Sage.

Nelson DL, and Cox MM. (2005). *Lehninger Principles of Biochemistry*. 7th ed. W. H. Freeman.

Newman MEJ. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* 46, 323-351. Available online at [http://tuvalu.santafe.edu/~aaronc/courses/7000/readings/Newman\\_05\\_PowerLawsParetoDistributionsAndZipfsLaw.pdf](http://tuvalu.santafe.edu/~aaronc/courses/7000/readings/Newman_05_PowerLawsParetoDistributionsAndZipfsLaw.pdf). Accessed 14 May 2018 by JKH.

Ryan KJ, and Ray CG, eds. (2004). *Sherris Medical Microbiology* (4th ed.). McGraw Hill.

Wolfram Research, Inc. (2022). *Mathematica Home Edition* v12.3.1.

White H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* 48, , pp. 817-838.