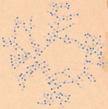
Generating Functions for Random Networks

Complex Networks
CSYS/MATH 303, Spring, 2011

Prof. Peter Dodds

Department of Mathematics & Statistics
Center for Complex Systems
Vermont Advanced Computing Center
University of Vermont

Outline



Generating functions

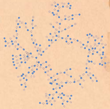
- Idea: Given a sequence $a_0, a_1, a_2, \ldots,$ associate each element with a distinct function or other mathematical object.
- Well-chosen functions allow us to manipulate sequences and retrieve sequence elements.

Definition:

- The generating function (g.f.) for a sequence $\{a_n\}$ is

$$F(x) = \sum_{n=0}^{\infty} a_n x^n.$$

- Roughly: transforms a vector in R^∞ into a function defined on R^1.
- Related to Fourier, Laplace, Mellin, ...



Simple example

Rolling dice:

- $p_k^{(\square)} = \mathbf{Pr}(\text{throwing a } k) = 1/6$ where $k = 1, 2, \ldots, 6$.
- $$F^{(\square)}(x) = \sum_{k=1}^{6} p_k x^k = \frac{1}{6}(x + x^2 + x^3 + x^4 + x^5 + x^6).$$
- We'll come back to this simple example as we derive various delicious properties of generating functions.

Example

- Take a degree distribution with exponential decay:

$$P_k = ce^{-\lambda k}$$

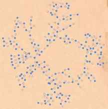
where $c = 1 - e^{-\lambda}$.

- The generating function for this distribution is

$$F(x) = \sum_{k=0}^{\infty} P_k x^k = \sum_{k=0}^{\infty} ce^{-\lambda k} x^k = \frac{c}{1 - xe^{-\lambda}}.$$

- Notice that $F(1) = c/(1 - e^{-\lambda}) = 1$.
- For probability distributions, we must always have $F(1) = 1$ since

$$F(1) = \sum_{k=0}^{\infty} P_k 1^k = \sum_{k=0}^{\infty} P_k = 1.$$



Properties of generating functions

- Average degree:

$$\langle k \rangle = \sum_{k=0}^{\infty} k P_k = \sum_{k=0}^{\infty} k P_k x^{k-1} \bigg|_{x=1}$$

$$= \frac{\mathrm{d}}{\mathrm{d}x} F(x) \bigg|_{x=1} = F'(1)$$

- In general, many calculations become simple, if a little abstract.
- For our exponential example:

$$F'(x) = \frac{(1 - e^{-\lambda}) e^{-\lambda}}{(1 - x e^{-\lambda})^2}.$$

- So:

$$\langle k \rangle = F'(1) = \frac{e^{-\lambda}}{(1 - e^{-\lambda})}.$$

Properties of generating functions

Useful pieces for probability distributions:

- Normalization:

$$F(1) = 1$$

- First moment:

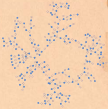
$$\langle k \rangle = F'(1)$$

- Higher moments:

$$\langle k^n \rangle = \left(x \frac{\mathrm{d}}{\mathrm{d}x} \right)^n F(x) \Big|_{x=1}$$

- kth element of sequence (general):

$$P_k = \frac{1}{k!} \frac{\mathrm{d}^k}{\mathrm{d}x^k} F(x) \Big|_{x=0}$$



Edge-degree distribution

- Recall our condition for a giant component:

$$\langle k \rangle_R = \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} > 1.$$

- Let's rëexpress our condition in terms of generating functions.
- We first need the g.f. for R_k.
- We'll now use this notation:

  $F_P(x)$ is the g.f. for P_k.
  $F_R(x)$ is the g.f. for R_k.

- Condition in terms of g.f. is:

$$\langle k \rangle_R = F_R'(1) > 1.$$

- Now find how F_R is related to F_P...

Edge-degree distribution

- We have

$$F_R(x) = \sum_{k=0}^{\infty} R_k x^k = \sum_{k=0}^{\infty} \frac{(k+1)P_{k+1}}{\langle k \rangle} x^k.$$

Shift index to $j = k + 1$ and pull out $\frac{1}{\langle k \rangle}$:

$$F_R(x) = \frac{1}{\langle k \rangle} \sum_{j=1}^{\infty} j P_j x^{j-1} = \frac{1}{\langle k \rangle} \sum_{j=1}^{\infty} P_j \frac{\mathrm{d}}{\mathrm{d}x} x^j$$

$$= \frac{1}{\langle k \rangle} \frac{\mathrm{d}}{\mathrm{d}x} \sum_{j=1}^{\infty} P_j x^j = \frac{1}{\langle k \rangle} \frac{\mathrm{d}}{\mathrm{d}x} \left(F_P(x) - P_0 \right) = \frac{1}{\langle k \rangle} F_P'(x).$$

Finally, since $\langle k \rangle = F_P'(1)$,

$$\boxed{F_R(x) = \frac{F_P'(x)}{F_P'(1)}}$$

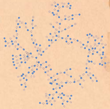
Edge-degree distribution

- Recall giant component condition is $\langle k \rangle_R = F'_R(1) > 1$.
- Since we have $F_R(x) = F'_P(x)/F'_P(1)$,

$$F'_R(x) = \frac{F''_P(x)}{F'_P(1)}.$$

- Setting $x = 1$, our condition becomes

$$\boxed{\frac{F''_P(1)}{F'_P(1)} > 1}$$



Size distributions

To figure out the size of the largest component (S_1), we need more resolution on component sizes.

Definitions:

- π_n = probability that a random node belongs to a finite component of size $n < \infty$.
- ρ_n = probability a random link leads to a finite subcomponent of size $n < \infty$.

Local-global connection:

$$P_k, R_k \Leftrightarrow \pi_n, \rho_n$$

neighbors $\Leftrightarrow$ components

Size distributions

G.f.'s for component size distributions:

- $$F_\pi(x) = \sum_{n=0}^{\infty} \pi_n x^n \text{ and } F_\rho(x) = \sum_{n=0}^{\infty} \rho_n x^n$$

The largest component:

- Subtle key: $F_\pi(1)$ is the probability that a node belongs to a finite component.
- Therefore: $S_1 = 1 - F_\pi(1)$.

Our mission, which we accept:

- Find the four generating functions

$$F_P, F_R, F_\pi, \text{ and } F_\rho.$$

Useful results we'll need for g.f.'s

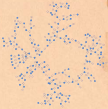
Sneaky Result 1:

- Consider two random variables U and V whose values may be $0, 1, 2, \ldots$
- Write probability distributions as U_k and V_k and g.f.'s as F_U and F_V.
- SR1: If a third random variable is defined as

$$W = \sum_{i=1}^{U} V^{(i)} \text{ with each } V^{(i)} \overset{d}{=} V$$

then

$$\boxed{F_W(x) = F_U\left(F_V(x)\right)}$$



Proof of SR1:

Write probability that variable W has value k as W_k.

$$W_k = \sum_{j=0}^{\infty} U_j \times \Pr(\text{sum of } j \text{ draws of variable } V = k)$$

$$= \sum_{j=0}^{\infty} U_j \sum_{\substack{\{i_1, i_2, \ldots, i_j\} | \\ i_1 + i_2 + \ldots + i_j = k}} V_{i_1} V_{i_2} \cdots V_{i_j}$$

$$\therefore F_W(x) = \sum_{k=0}^{\infty} W_k x^k = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} U_j \sum_{\substack{\{i_1, i_2, \ldots, i_j\} | \\ i_1 + i_2 + \ldots + i_j = k}} V_{i_1} V_{i_2} \cdots V_{i_j} x^k$$

$$= \sum_{j=0}^{\infty} U_j \sum_{k=0}^{\infty} \sum_{\substack{\{i_1, i_2, \ldots, i_j\} | \\ i_1 + i_2 + \ldots + i_j = k}} V_{i_1} x^{i_1} V_{i_2} x^{i_2} \cdots V_{i_j} x^{i_j}$$

Proof of SR1:

With some concentration, observe:

$$F_W(x) = \sum_{j=0}^{\infty} U_j \sum_{k=0}^{\infty} \underbrace{\underbrace{\sum_{\substack{\{i_1,i_2,\ldots,i_j\}| \\ i_1+i_2+\ldots+i_j=k}} V_{i_1} x^{i_1} V_{i_2} x^{i_2} \cdots V_{i_j} x^{i_j}}_{x^k \text{ piece of } \left(\sum_{i'=0}^{\infty} V_{i'} x^{i'}\right)^j}}_{\left(\sum_{i'=0}^{\infty} V_{i'} x^{i'}\right)^j = (F_V(x))^j}$$

$$= \sum_{j=0}^{\infty} U_j (F_V(x))^j$$

$$= F_U(F_V(x)) \checkmark$$

Useful results we'll need for g.f.'s

Sneaky Result 2:

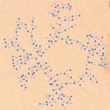
- Start with a random variable U with distribution U_k $(k = 0, 1, 2, \ldots)$
- SR2: If a second random variable is defined as

$$V = U + 1 \text{ then } \boxed{F_V(x) = xF_U(x)}$$

- Reason: $V_k = U_{k-1}$ for $k \geq 1$ and $V_0 = 0$.
-

$$\therefore F_V(x) = \sum_{k=0}^{\infty} V_k x^k = \sum_{k=1}^{\infty} U_{k-1} x^k$$

$$= x \sum_{j=0}^{\infty} U_j x^j = xF_U(x). \checkmark$$



Useful results we'll need for g.f.'s

Generalization of SR2:

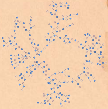
- (1) If $V = U + i$ then

$$F_V(x) = x^i F_U(x).$$

- (2) If $V = U - i$ then

$$F_V(x) = x^{-i} F_U(x)$$

$$= x^{-i} \sum_{k=0}^{\infty} U_k x^k$$

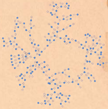


Connecting generating functions

- Goal: figure out forms of the component generating functions, F_π and F_ρ.
- π_n = probability that a random node belongs to a finite component of size n

$$= \sum_{k=0}^{\infty} P_k \times \Pr\left(\begin{array}{c} \text{sum of sizes of subcomponents} \\ \text{at end of } k \text{ random links} = n-1 \end{array} \right)$$

- Therefore: $$F_\pi(x) = \underbrace{x}_{\text{SR2}} \underbrace{F_P\left(F_\rho(x)\right)}_{\text{SR1}}$$
- Extra factor of x accounts for random node itself.

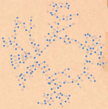


Connecting generating functions

- ρ_n = probability that a random link leads to a finite subcomponent of size n.
- Invoke one step of recursion: ρ_n = probability that in following a random edge, the outgoing edges of the node reached lead to finite subcomponents of combined size $n-1$,

$$= \sum_{k=0}^{\infty} R_k \times \Pr\left(\begin{array}{l}\text{sum of sizes of subcomponents}\\ \text{at end of } k \text{ random links} = n-1\end{array}\right)$$

- Therefore: $$\boxed{F_\rho(x) = \underbrace{x}_{\text{SR2}}\ \underbrace{F_R\left(F_\rho(x)\right)}_{\text{SR1}}}$$
- Again, extra factor of x accounts for random node itself.



Connecting generating functions

- We now have two functional equations connecting our generating functions:

$$F_\pi(x) = xF_P\left(F_\rho(x)\right) \quad \text{and} \quad F_\rho(x) = xF_R\left(F_\rho(x)\right)$$

- Taking stock: We know $F_P(x)$ and $F_R(x) = F_P'(x)/F_P'(1)$.
- We first untangle the second equation to find F_ρ
- We can do this because it only involves F_ρ and F_R.
- The first equation then immediately gives us F_π in terms of F_ρ and F_R.

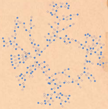
Component sizes

- Remembering vaguely what we are doing:

  Finding F_π to obtain the fractional size of the largest component $S_1 = 1 - F_\pi(1)$.
- Set $x = 1$ in our two equations:

$$F_\pi(1) = F_P\left(F_\rho(1)\right) \quad \text{and} \quad F_\rho(1) = F_R\left(F_\rho(1)\right)$$

- Solve second equation numerically for $F_\rho(1)$.
- Plug $F_\rho(1)$ into first equation to obtain $F_\pi(1)$.



Component sizes

Example: Standard random graphs.

- We can show $F_P(x) = e^{-\langle k \rangle (1-x)}$

$$\therefore F_R(x) = F_P'(x)/F_P'(1) = e^{-\langle k \rangle (1-x)} / e^{-\langle k \rangle (1-x')}|_{x'=1}$$

$$= e^{-\langle k \rangle (1-x)} = F_P(x)$$ ...aha!

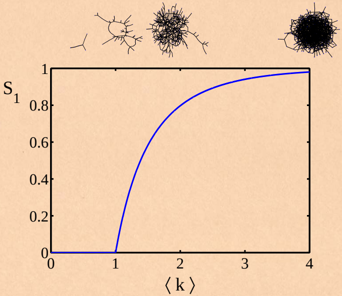
- RHS's of our two equations are the same.
- So $F_\pi(x) = F_\rho(x) = xF_R(F_\rho(x)) = xF_R(F_\pi(x))$
- Why our dirty (but wrong) trick worked earlier...

Component sizes

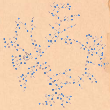
- We are down to $F_\pi(x) = xF_R(F_\pi(x))$ and $F_R(x) = e^{-\langle k \rangle (1-x)}$.
-
$$\therefore F_\pi(x) = xe^{-\langle k \rangle (1-F_\pi(x))}$$
- We're first after $S_1 = 1 - F_\pi(1)$ so set $x = 1$ and replace $F_\pi(1)$ by $1 - S_1$:

$$1 - S_1 = e^{-\langle k \rangle S_1}$$

$$\text{Or: } \langle k \rangle = \frac{1}{S_1} \ln \frac{1}{1 - S_1}$$



- Just as we found with our dirty trick ...
- Again, we (usually) have to resort to numerics ...



Average component size

- Next: find average size of finite components $\langle n \rangle$.
- Using standard G.F. result: $\langle n \rangle = F'_\pi(1)$.
- Try to avoid finding $F_\pi(x)$...
- Starting from $F_\pi(x) = xF_P\left(F_\rho(x)\right)$, we differentiate:

$$F'_\pi(x) = F_P\left(F_\rho(x)\right) + xF'_\rho(x)F'_P\left(F_\rho(x)\right)$$

- While $F_\rho(x) = xF_R\left(F_\rho(x)\right)$ gives

$$F'_\rho(x) = F_R\left(F_\rho(x)\right) + xF'_\rho(x)F'_R\left(F_\rho(x)\right)$$

- Now set $x = 1$ in both equations.
- We solve the second equation for $F'_\rho(1)$ (we must already have $F_\rho(1)$).
- Plug $F'_\rho(1)$ and $F_\rho(1)$ into first equation to find $F'_\pi(1)$.

Average component size

Example: Standard random graphs.

- Use fact that $F_P = F_R$ and $F_\pi = F_\rho$.
- Two differentiated equations reduce to only one:

$$F'_\pi(x) = F_P\left(F_\pi(x)\right) + xF'_\pi(x)F'_P\left(F_\pi(x)\right)$$

$$\text{Rearrange:} \quad F'_\pi(x) = \frac{F_P\left(F_\pi(x)\right)}{1 - xF'_P\left(F_\pi(x)\right)}$$

- Simplify denominator using $F'_P(x) = \langle k \rangle F_P(x)$
- Replace $F_P(F_\pi(x))$ using $F_\pi(x) = xF_P(F_\pi(x))$.
- Set $x = 1$ and replace $F_\pi(1)$ with $1 - S_1$.

$$\text{End result:} \quad \langle n \rangle = F'_\pi(1) = \frac{(1 - S_1)}{1 - \langle k \rangle (1 - S_1)}$$

Average component size

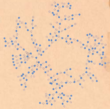
- Our result for standard random networks:

$$\langle n \rangle = F'_{\pi}(1) = \frac{(1 - S_1)}{1 - \langle k \rangle (1 - S_1)}$$

- Recall that $\langle k \rangle = 1$ is the critical value of average degree for standard random networks.
- Look at what happens when we increase $\langle k \rangle$ to 1 from below.
- We have $S_1 = 0$ for all $\langle k \rangle < 1$ so

$$\langle n \rangle = \frac{1}{1 - \langle k \rangle}$$

- This blows up as $\langle k \rangle \to 1$.
- Reason: we have a power law distribution of component sizes at $\langle k \rangle = 1$.
- Typical critical point behavior....



Average component size

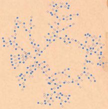
- Limits of $\langle k \rangle = 0$ and ∞ make sense for

$$\langle n \rangle = F'_{\pi}(1) = \frac{(1 - S_1)}{1 - \langle k \rangle (1 - S_1)}$$

- As $\langle k \rangle \to 0$, $S_1 = 0$, and $\langle n \rangle \to 1$.
- All nodes are isolated.
- As $\langle k \rangle \to \infty$, $S_1 \to 1$ and $\langle n \rangle \to 0$.
- No nodes are outside of the giant component.

Extra on largest component size:

- For $\langle k \rangle = 1$, $S_1 \sim N^{2/3}$.
- For $\langle k \rangle < 1$, $S_1 \sim \log N$.



References I