Four generating functions from previous lecture

Deg: \( g_0, g_1 \)

Size: \( h_0, h_1 \)

\[
\begin{align*}
  h_0'(1) &= 1 + g_0'(1)h_1'(1) \quad \text{(I)} \\
  h_1'(1) &= 1 + g_1'(1)h_1'(1) \quad \text{(II)}
\end{align*}
\]

From (II), we have

\[
  h_1'(1) = \frac{1}{1 + g_1'(1)} \quad \text{(III)}
\]

Plugging (III) back to (I)

\[
  h_0'(1) = 1 + \frac{g_0'(1)}{1 - g_1'(1)}
\]

Therefore, point at which the giant components appears when \( g_1'(1) = 1 \)

We also know that:

\[
  g_1'(1) = g_0''(1) / g_0'(1) = 1 \quad \rightarrow \quad g_0''(1) = g_0'(1)
\]

By definition:

\[
\begin{align*}
  g_0(x) &= \sum P_k x^k \\
  g_0'(x) &= \sum k P_k x^{k-1} \\
  g_0''(x) &= \sum k(k-1) P_k x^{k-2}
\end{align*}
\]

Since \( g_0''(1) = g_0'(1) \), let \( x = 1 \)

\[
\begin{align*}
  \sum k P_k &= \sum k(k-1) P_k \\
  \sum (k^2-2k) P_k &= 0 \\
  \sum k(k-2) P_k &= 0
\end{align*}
\]

This is the end of random graph materials.

**High Dimensional Data**

10⁶ papers
25,000 dimensional vector to represent the words used in the papers

This is a very large dimension, so we can randomly select a 300 dimensional subspace and project all the data points to such subspace. In fact, we can pick a subspace such that we minimize the error between data and its projection.

**SVD: Singular Value Decomposition**

Let \( A \) be the 10⁶×25000 matrix of data from your set of papers. Any matrix \( A \) can be rewritten as \( A = U \Sigma V^T \) where \( U \) and \( V \) are orthonormal and \( \Sigma \) is diagonalised:
\[ \Sigma = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_N \end{pmatrix} \text{ with } \sigma_1 \geq \sigma_2 \geq \sigma_3 \geq \ldots \geq \sigma_N \]

Create a new matrix \( \Sigma^{(k)} \) by keeping only the \( k \) largest \( \sigma \)'s in \( \Sigma \):

\[ \Sigma^{(k)} = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \]

Then write \( A^{(k)} = U \Sigma^{(k)} V^T \).

What is the error between \( A \) and \( A^{(k)} \)?

**Norms**

- Frobenius norm: \( |A|_F^2 = \sum_{i,j} a_{ij}^2 = \text{sum of squares of all of the elements of } A \).

  Let \( a_i \) be the \( i \)th column of \( A \). Then \( (A^T A)_{ij} = (a_i)^T a_j \).

  So \( \text{Tr}(A^T A) = \sum (a_i)^T a_i = \sum |a_i|^2 = \text{sum of squares of elements of } A = |A|_F^2 \)

- 2-norm:

  For a vector \( x = (x_1, x_2, \ldots, x_n) \), \( |x|_2 = \sqrt{x_1 + x_2 + x_3 + \ldots} \)

  For a matrix, \( |A|_2 = \max_{|x|=1} |Ax|_2 \)

- It turns out that the 2-norm of \( A \) is the maximum eigenvalue of \( A \), and the Frobenius norm squared is the sum of the squares of the eigenvalues. To see this, take the 2-norm and Frobenius norm of a diagonalised matrix and note the 2nd lemma below.

**Lemma**: \( |AB|_2 \leq |A|_2 |B|_2 \)

**Pf**: Let \( y \) be the value of \( x \) that maximises \( |A|_2 = \max_{|x|=1} |Ax|_2 \).

Let \( z = By \).

Then \( |AB|_2 = |Az|_2 = \frac{|z|_2}{|z|_2} |A|_2 |z|_2 \leq |A|_2 |B|_2 \) since \( |z|_2 \leq |B|_2 \)

**Lemma**: Let \( Q \) be an orthonormal matrix. For all \( x \), \( |Qx|_2 = |x|_2 \).

**Pf**: \( |Qx|_2^2 = x^T Q^T Q x = x^T x = |x|_2^2 \) because \( Q^T = Q^{-1} \), so \( Q^T Q = I \).

**Lemma**: Let \( Q \) be an orthonormal matrix. For all \( A \), \( |QA|_2 = |A|_2 \).

**Pf**: This follows directly from the previous lemma.

These lemmas also hold for the Frobenius norm as well.

It is straightforward from this to show that for \( A = U \Sigma V^T \), \( |A| = |\Sigma| \) for both norms.

**Thm**: \( |A - A^{(k)}|_2 \leq \min_{\text{rank}(B) \leq k} |A - B| \). In other words, \( A^{(k)} \) is the best rank-\( k \) approximation to \( A \) in terms of total error (but not necessarily error of a given element, so this is not always “best” in every application). Proof next lecture. This is easy to prove for the 2-norm but very difficult for the Frobenius norm.