CSI 445/660 – Part 7

(Models for Random Graphs)

Models for Random Graphs

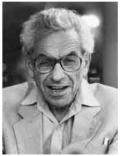
References:

1. R. Albert and A. Barabasi, "Statistical Mechanics of Complex Networks", *Reviews of Modern Physics*, Vol. 74, Jan. 2002, pp. 47–97.
2. Chapter 18 of [EK].

Motivation:

- Provide methods for generating large random networks.
- Such synthetic networks are useful in
  - testing applications and
  - checking whether or not a given social network is similar to a random network.
- Many methods have been proposed; each is useful in certain applications.

Erdős-Rényi-Gilbert Model



- Paul Erdős (1913–1996)
- See slides for Part 1 for additional information.

- Alfréd Rényi (1921–1970)
- Ph.D., University of Szeged, 1947.
- Many contributions to Mathematics.

- Edger Gilbert (1923–2013)
- Ph.D., MIT, 1948.
- Worked on Coding Theory at Bell Labs, NJ.

Erdős-Rényi-Gilbert Model (continued)

Basic information about the model:

- Proposed by Gilbert and developed extensively by Erdős and Rényi.
- Commonly known as the Erdős-Rényi (ER) model.
- Uses two parameters:
  1. the number of nodes (n) and
  2. the probability (p) of an edge between any pair of nodes.
- Also called the $G(n, p)$ model.
- Usually, p is a function of n (e.g. $p = 1/n$).
- Edges between pairs of nodes are chosen **independently**.

Erdős-Rényi-Gilbert Model (continued)

Note: Assume that the nodes are numbered 1 through n.

Algorithm for ER model graph generation:

for $i = 1$ **to** $n - 1$ **do** $\{$
    for $j = i + 1$ **to** n **do** $\{$
        Add edge $\{i, j\}$ with probability p.
    $\}$
$\}$

Notes:

- The above algorithm generates an undirected graphs.
- Can be easily modified to generate directed graphs.
- We will restrict our attention to undirected graphs.

Some Random Graph Generation Facilities in CINET

- **$G(n, p)$ random graph**: This generates a random graph under the ER model.
- **$G(n, p)$ component**: This generates a random graph under the ER model and gives the distribution of the sizes of the connected components (in the form of a table).
- **$G(n, m)$ random graph**: This generates a random graph with n nodes and m edges.
- **(n, d)-random regular graph**:
  - A graph is **regular** if every node has the same degree.
  - This generator produces a random graph with n nodes where each node has degree $= d$.

Some Random Graph ... CINET (continued)

- $G(n, r)$ **random graph**: This generates a random **geometric** graph as follows:
  - A total of n points are randomly chosen within the unit cube.
  - Each point is a node of the graph.
  - An edge is added between a pair of nodes if the distance between the corresponding pair of points is at most r.
  - Such graphs arise in the study of wireless (ad hoc) networks.

Some simple properties:

1. Expected degree of any node $= p(n-1)$.

Proof: Consider any node v.

- Node v may have up to $n-1$ possible edges, say $e_1, e_2, \ldots, e_{n-1}$, to the other nodes.
- Let X_i be a RV associated with edge e_i, $1 \le i \le n-1$: $X_i = 1$ if edge e_i is present and 0 otherwise. (X_i is called an **indicator** RV.)
- $\mathrm{Degree}(v) = X_1 + X_2 + \ldots + X_{n-1}$ is another RV.
- Now, $\Pr\{X_i = 1\} = p$ and $\Pr\{X_i = 0\} = 1 - p$.
  So, $\mathrm{E}[X_i] = p \;\; (1 \le i \le n-1)$.
- So, by linearity of expectation, $\mathrm{E}[\mathrm{Degree}(v)] = p(n-1)$.

Erdős-Rényi-Gilbert Model (continued)

Some simple properties (continued):

2. Expected number of edges $= n(n-1)p/2$.

Proof:

- Introduce an indicator RV Y_i for each of the $N = n(n-1)/2$ possible edges.
- Let Y denote the RV for the number of edges. Thus,

$$Y = Y_1 + Y_2 + \ldots + Y_N.$$

- As before, $\mathrm{E}[Y_i] = p, \ (1 \leq i \leq N)$.
- By linearity of expectation, $\mathrm{E}[Y] = pN = pn(n-1)/2$.

Some simple properties (continued):

3. Let $\pi_k(v)$ denote the probability that node v has degree $= k$ $(0 \leq k \leq n-1)$. Then,

$$\pi_k(v) \;=\; \binom{n-1}{k} p^k (1-p)^{n-1-k}.$$

- This called the **binomial distribution**.
- This is the same probability as getting k heads from $n-1$ tosses of a coin, where the probability of heads $= p$.

Erdős-Rényi-Gilbert Model (continued)

Some non-trivial properties: The following results due to Erdős and Rényi are **asymptotic** (i.e., they hold for large n).

Condition	**Property of $G(n, p)$**
$p < 1/n$	Almost surely has **no** connected component of size larger than $c_1 \log_2 n$ for some constant c_1.
$p = 1/n$	Almost surely has a giant component of size at least $c_2\, n^{2/3}$ for some constant c_2.
$p > 1/n$	Almost surely has a giant component of size at least αn for some constant α $(0 < \alpha < 1)$. All other components will almost surely have size $\leq \beta\, log_2 n$ for some constant β.
$p = 1/2$	With high probability, the size of the largest clique is $\approx 2 \log_2 n$.

ER Model and the Web Graph

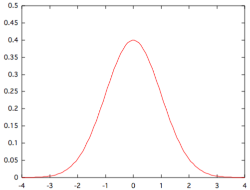
Is the ER model appropriate for the web graph?

- Consider the node degrees as n increases.
- Each **edge**: A random variable (RV), which has the value 1 with probability p and the value 0 with probability $1 - p$.
- For any node v, degree(v) is the **sum** of the $n-1$ of the edge RVs.
- These $n-1$ RVs are **independent** and **identically distributed** (iid).

Central Limit Theorem (simplified statement):

As $n \to \infty$, the sum of n iid RVs approaches the **normal** (or Gaussian) distribution.

ER Model and the Web Graph (continued)



Note: For such a distribution and large values of k, the fraction of nodes with degree k can be shown to **decrease exponentially** (i.e., something like 2^{-k}).

Experimental evidence: The fraction of nodes with degree k in the web graph decreases (roughly) as $1/k^2$.

Comparison: Suppose $k = 1000$. Then $1/k^2 = 10^{-6}$. However,

$$2^{-k} = 1/2^{1000} < 10^{-250}$$

which is much smaller than 10^{-6}.

- So, ER model is **not** appropriate for the web graph.
- A more appropriate model is that of **power law** (or **scale-free**) graphs.

Definition of Power Law

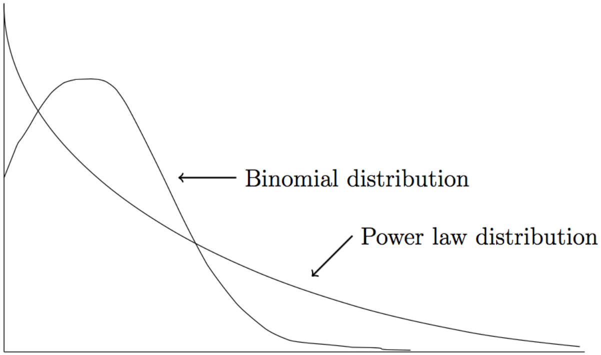
Definition: A function $f(k)$ exhibits **power law** behavior if it decreases with k as k^{-c} for some positive constant c.

Examples from empirical studies: (from Chapter 18 of [EK] text)

- The fraction of telephone numbers that receive k calls per day is roughly proportional to $1/k^2$.
- The fraction of books bought by k people is roughly proportional to $1/k^3$.
- The fraction of scientific papers that receive k citations is roughly proportional to $1/k^3$.

Note: Many measures of popularity seem to exhibit power law behaviors.

A Characteristic of Power Law Distribution



Note: Power law distribution has a **heavy tail**.

How to Check for Power Law

Given: The values of function $f(k)$ for different values of k.

k	$f(k)$
1.0	445.7
1.5	411.3
$\vdots$	$\vdots$
31.2	13.9

- We want to check whether the data exhibits a power law behavior.
- If so, we want to find the exponent c.

Idea: Suppose the data exhibits power law behavior; that is,

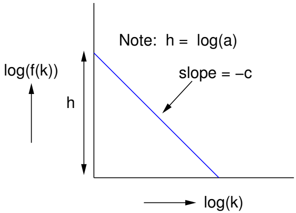
$$f(k) \;=\; a \times k^{-c} \quad \text{for some constants } a \text{ and } c.$$

Then

$$\log_{10}(f(k)) \;=\; \log_{10}(a) - c\log_{10}(k).$$

Observation: If $\log_{10}(f(k))$ is plotted against $\log_{10}(k)$, the graph will be a straight line.

How to Check for Power Law (continued)



- Slope of the line $= -c$.
- y-intercept of the line $= \log_{10}(a)$.

Note: Many plotting programs can produce log-log plots.

Computing the exponent:

- Consider the function values $f(k_1)$ and $f(k_2)$ at two values k_1 and k_2.
- Let $x_1 = \log_{10}(k_1)$ and $x_2 = \log_{10}(k_2)$.
- Let $y_1 = \log_{10}(f(k_1))$ and $y_2 = \log_{10}(f(k_2))$.
- Slope of the line $= (y_2 - y_1)/(x_2 - x_1)$ and the power law exponent $c = -$slope.

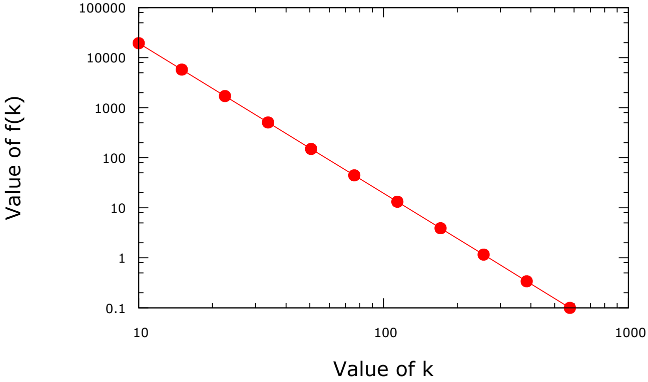
How to Check for Power Law (continued)

Problem: Check whether the data shown in the following table exhibits power law behavior; if so, find the power law exponent.

k	$f(k)$	k	$f(k)$
10.00	19500.00	113.91	13.19
15.00	5777.78	170.86	3.91
22.50	1711.93	256.29	1.16
33.75	507.24	384.43	0.34
50.62	150.29	576.65	0.10
75.94	44.53		

Solution: The log-log plot for this data is shown on the next slide.

Log-Log plot for the data:



Note: Since the log-log plot is a straight line, the given data exhibits power law behavior.

How to Check for Power Law (continued)

Value of the power law exponent:

- From the given data set choose $k_1 = 22.50$ and $k_2 = 33.75$.
  So, $x_1 = \log_{10}(22.50)$ and $x_2 = \log_{10}(33.75)$.

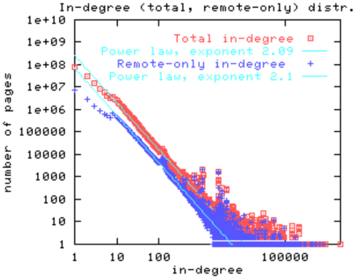
- Also from the given data set, $f(k_1) = 1711.93$ and $f(k_2) = 507.24$.
  So, $y_1 = \log_{10}(1711.93)$ and $y_2 = \log_{10}(507.24)$.

- Slope $= (y_2 - y_1)/(x_2 - x1) = -2.9999$.

- So, power law exponent $= 2.9999$ (which is close to 3.0).

Power Law Example: Web Graph

An example from [EK]:



- From **[Broder et al. 2000]**.
- Shows both total indegree (**red**) and remote-only indegree (**blue**).
- The corresponding power law exponents are (approximately) 2.09 and 2.1 respectively.

- The power law behavior of the web graph suggests that its evolution **cannot** be captured by the ER model.
- **Question:** Which random graph model allows node degrees to have a power law distribution?
- **Answer:** The **preferential attachment** (or "rich get richer") model.

Preferential Attachment Model

- Herbert A. Simon (1916–2001)
- Ph.D. (Political Science), University of Chicago, 1943.
- Taught at Carnegie Mellon University.

- Contributed to many areas (e.g. Political Science, Economics, Psychology, Cognitive Science, Computer Science).
- Won the Nobel Prize in Economics (for his contributions to decision-making processes in organizations).
- Also won the Turing Award in Computer Science (for his contributions to AI).

Preferential Attachment Model (continued)

- Simon **[Biometrika, 1955]** developed a general model to explain power law behavior in many different situations.

  Example: The fraction of cities with with population k was known to follow a power law.

- Simon's model allowed the derivation of the corresponding power law using the following assumption:

  The **rate** *at which the population of a city grows is proportional to the* **current size** *of the population.*

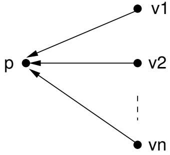
- Hence the name "rich get richer" model.
- The name "preferential attachment" was coined later (by Albert & Barabasi).

Preferential Attachment and the Web Graph

Web graph:

- Directed graph.
- Nodes are web pages; the directed edge (x, y) means that that web page x has a link to web page y.
- Indegrees exhibit a power law behavior.
- Interpretation of "rich get richer" idea:

  Popular web pages are likely to get more in-links, further increasing their popularity.



- **Consequence:** Web pages with large **indegrees** exist.

Generating a Directed Graph with Power Law Behavior

Goal: To generate a random **directed** graph where **indegrees** have a power law behavior.

Assumptions:

- There are n web pages (numbered 1 through n) and they arrive one at a time.
- A probability value p, $0 < p < 1$, which provides an indication of the likelihood of preferential attachment, is given.

  Note: The value of p determines the power law exponent.
- Each node has an outdegree of 1.

  Note: The graph generation procedure can be generalized to remove this assumption.

Description of the Algorithm: See Handout 7.1.

Generating an Undirected Graph with Power Law Behavior

Goal: To generate a random **undirected** graph where node **degrees** have a power law behavior.

Ref: **[Albert & Barabasi, 2002]**

Assumptions:

- Initially, there are $m_0 \geq 1$ nodes (numbered 1 through m_0). (When the algorithm ends, there are n nodes, numbered 1 through n.)
- For each new node, $m \leq m_0$ edges are added.
- In the resulting undirected graph, degrees follow a power law with exponent $c \approx 3$.

Description of the Algorithm: See Handout 7.2.

Note: CINET provides a graph generator for this model.

Example for Step 1(i) of the Algorithm in Handout 7.2

Note: Step 1(i) of the algorithm implements the "rich get richer" idea.

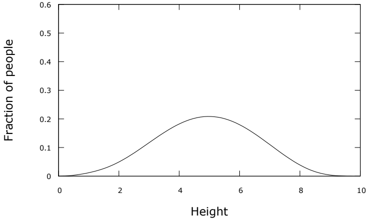
Example:

- Let $m = 1$; that is, each new node will get one edge.
- There are 4 nodes (numbered 1, 2, 3 and 4) and the new one is node 5.
- Let the degrees of nodes 1, 2, 3 and 4 be 3, 3, 2 and 2 respectively.
- Current sum of degrees $= 3 + 3 + 2 + 2 \; = \; 10$.
- For node 5:
  - $\Pr\{\text{Edge to node 1}\} \; = \; 3/10$.
  - $\Pr\{\text{Edge to node 2}\} \; = \; 3/10$.
  - $\Pr\{\text{Edge to node 3}\} \; = \; 2/10$.
  - $\Pr\{\text{Edge to node 4}\} \; = \; 2/10$.

A Note on Scale-Free Graphs

- The terms "power law graphs" and "scale-free graphs" are treated as synonyms in the literature.
- There are several interpretations of the phrase "scale-free".

Interpretation 1: (due to Albert & Barabasi)



- There is no person with a height of 9 feet or more; that is, at "higher scales", the proportion drops to zero.
- For power law graphs, the proportion is positive even for very large degrees; that is, there are nodes at "all scales".

A Note on Scale-Free Graphs (continued)

Interpretation 2: Let $P(d)$ denote the proportion of nodes with degree d.

- When $P(d)$ obeys a power law,

$$P(d) \;=\; \alpha\, d^{\beta}, \text{ for some } \alpha > 0 \text{ and } \beta < 0.$$

- For degree values d_1 and d_2,

$$\frac{P(d_1)}{P(d_2)} \;=\; \left(\frac{d_1}{d_2}\right)^{\beta}.$$

- Suppose we “scale” the degrees d_1 and d_2 by a factor k. Then,

$$\frac{P(k\,d_1)}{P(k\,d_2)} \;=\; \left(\frac{d_1}{d_2}\right)^{\beta} \;=\; \frac{P(d_1)}{P(d_2)}.$$

- So, the **ratio doesn’t change** when degrees are scaled; in this sense, power law graphs are “scale-free”.

A Note on Scale-Free Graphs (continued)

Interpretation 3: (due to Fan Chung & Linyuan Lu)

- The word "scale" is with respect to **time**.
- **Example:** Consider the algorithm for generating directed graphs with power law distribution.
  - At each time step, one new node and one directed edge are added.
  - Instead, consider a time interval of length t: t nodes arrive during the interval and t edges are added.
  - The power law exponent is **independent** of the value of t; thus, it is **free from any scaling with respect to time**.

Chung-Lu Model of Random Graphs

- Proposed by Fan Chung (University of California, San Diego) and Linyuan Lu (University of South Carolina).
- Generalizes the ER model.
- **Inputs:**
  - Integer n, the number of nodes.
  - A sequence of n non-negative numbers $\langle w_1, w_2, \ldots, w_n \rangle$ (called a **degree sequence**) such that

$$\max_{1 \leq i \leq n} \{w_i^2\} < \sum_{i=1}^{n} w_i .$$

- **Output:** A random graph with n nodes (numbered 1 through n) such that the **expected degree** of node i is w_i, $1 \leq i \leq n$.
- The graph may have **self loops**.

Description of the Algorithm: See Handout 7.3.

Chung-Lu Model (continued)

Properties of the Chung-Lu Model:

- Generalizes the ER model:
  - Let $w_i = np$, $1 \leq i \leq n$, where n and p are the parameters of the ER model.
  - Then, the probability of adding any edge $\{i, j\}$ is exactly p.
- Can also generate graphs where degrees satisfy a power law.
  - For a power law exponent $\beta \geq 2$, the weights are chosen as follows:

$$w_i = (i/nB)^{-\frac{1}{\beta-1}}, \quad 1 \leq i \leq n,$$

    where

$$B = \frac{1}{(\beta - 1)\xi(\beta)} \quad \text{and} \quad \xi(\beta) = \sum_{k=1}^{\infty} k^{-\beta}.$$

Chung-Lu Model (continued)

Properties of the Chung-Lu Model (continued):

- For $\beta > 3$:
  - The diameter of the resulting graph is $O(\log n)$ with high probability.
  - The average distance between any pair of nodes is $O(\log n / \log \log n)$ with high probability.
- Thus, small-world networks can also be generated using the Chung-Lu model.

Watts-Strogatz Model

- Proposed in 1998 by Duncan Watts (Yahoo Research) and Steven Strogatz (Cornell University).
- Predates preferential attachment models.
- Addresses two aspects which are **not** present in the ER model.
  - ER model does not generate an adequate number of hubs (i.e., high degree nodes).
  - The average clustering coefficient is small under the ER model.
- Watts & Strogatz also wanted the graphs to have a small diameter (i.e., the "small world" property).

Watts-Strogatz Model (continued)

Rewiring:

- Steps needed to "rewire" edge $\{c, d\}$ in the graph on the left.
  1. Delete edge $\{c, d\}$.
  2. Add an edge from c to some other node **without** causing multi-edges or self-loops.

- In the above example, edge $\{c, d\}$ may get replaced by $\{c, a\}$ or $\{c, e\}$, each with probability $= 1/2$.

- The graph with edge $\{c, d\}$ replaced by $\{c, a\}$ is shown on the right.
- Rewiring can decrease the average distance (by adding "long range" edges).

Watts-Strogatz Model (continued)

Inputs:

- The number of nodes: n.
- An even integer K, the average node degree in the resulting graph.
- The rewiring probability β.
- **Assumption:** $n \gg K \gg \ln n \gg 1$.

Output: An undirected graph with the following properties.

- The graph has n nodes and $nK/2$ edges. (Thus, the average node degree is K.)
- With high probability, the average distance between any pair of nodes is $\ln(n)/\ln(K)$.

Description of the Algorithm: See Handout 7.4.

Watts-Strogatz Model (continued)

Notes:

- If $\beta = 0$, there is no rewiring and the diameter remains large.
- If $\beta = 1$, every edge gets rewired; it is known that such graphs are similar to graphs under the ER model.
- If $C(0)$ represents the average clustering coefficient of the initial graph, empirical evidence suggests that the average clustering coefficient $C(\beta)$ after rewiring is given by

  $$C(\beta) \;=\; C(0)\,(1-\beta)^3 \;.$$

  If β is small, the clustering coefficient does not decrease much due to rewriting.

Watts-Strogatz Model (continued)

Limitations:

- Degree distribution does not correspond to that of common social networks.
- The value of n must be known. So, the model is not useful in generating graphs that evolve over time.

Final Remarks:

- Researchers have tried the rewiring approach starting from other initial graphs (e.g. grids).
- **Newman-Watts Model:** Instead of rewiring, add edges between randomly chosen pairs of nodes with with probability $= \beta$.
  - This version is easier to implement.
  - The resulting model has properties similar to the Watts-Strogatz model.

Review of Some Concepts Related to Probability

Discrete Random Variable

Basic Information:

- **Abbreviation:** RV for "random variable".
- A **discrete** RV X takes on values from a discrete set S.
- For each element $a \in S$, the probability that X takes on the value a is denoted by $\Pr\{X = a\}$.
- Note that $\sum_{a \in S} \Pr\{X = a\} \; = \; 1$.

Example 1: Suppose X is an RV representing the outcome of tossing a fair coin. Here, $S = \{T, H\}$ and $\Pr\{X = T\} = \Pr\{X = H\} = 1/2$. (Thus, both the values of X are **equally likely**.)

Example 2: Suppose Y is an RV representing the outcome of tossing a fair die. Here, $S = \{1, 2, 3, 4, 5, 6\}$ and $\Pr\{Y = i\} = 1/6$, for $1 \leq i \leq 6$. (Here, all the six values of Y are equally likely.)

Expectation of a Discrete RV

Expectation: If X is a discrete RV taking values over a set S of numbers, then the **expectation** of X, denoted by $\mathrm{E}[X]$, is defined by

$$\mathrm{E}[X] \;=\; \sum_{a \in S} a \times \Pr\{X = a\}$$

Example 1: Suppose Y is an RV representing the outcome of tossing a fair die. Here, $S = \{1, 2, 3, 4, 5, 6\}$ and $\Pr\{Y = i\} = 1/6$, for $1 \leq i \leq 6$. Then,

$$\mathrm{E}[Y] \;=\; \sum_{i=1}^{6} i/6 \;=\; 3.5$$

Note: When all the values in S are equally likely, the expectation is equal to **average** (or **mean** value).

Expectation of a Discrete RV

Example 2: Suppose Z is an RV representing the outcome of tossing a **loaded** die. Again, $S = \{1, 2, 3, 4, 5, 6\}$. Let $\Pr\{Z = 1\} = 1/2$ and $\Pr\{Z = i\} = 1/10$, for $2 \leq i \leq 6$. Then,

$$\mathrm{E}[Z] \; = \; 1 \times 1/2 \; + \; \sum_{i=2}^{6} i/10 \; = \; 2.5$$

Linearity of Expectation: Suppose X_1, X_1, ..., X_n are RVs and a new RV X is defined by

$$X \; = \; X_1 + X_2 + \ldots + X_n \; .$$

Then

$$\mathrm{E}[X] \; = \; \mathrm{E}[X_1] + \mathrm{E}[X_2] + \ldots + \mathrm{E}[X_n].$$

Note: The above equation holds **even if there are dependencies among the RVs**.

Application of Linearity of Expectation:

Problem: Suppose we throw **two** fair dice. Find the expectation of the sum of the face values of the two dice.

Solution: Let W denote the RV that represents the sum of the face values of the two dice.

Method I (somewhat tedious): The possible values for the RV W are $\{2, 3, 4, \ldots, 12\}$. We first compute the probability of each these possible values.

$$\begin{aligned} \Pr\{W = 2\} &= 1/36 \\ \Pr\{W = 3\} &= 2/36 \\ &\vdots \\ \Pr\{W = 12\} &= 1/36 \end{aligned}$$

Then, we compute $\mathrm{E}[W]$ using the above values.

Expectation of a Discrete RV (continued)

Application of Linearity of Expectation (continued):

Method II: Let Y_1 and Y_2 denote the RVs corresponding to the face values of the two dice. Define a new RV $Y = Y_1 + Y_2$. Our goal is to compute $\mathrm{E}[Y]$.

By linearity of expectation, $\mathrm{E}[Y] = \mathrm{E}[Y_1] + \mathrm{E}[Y_2]$. As shown previously, $\mathrm{E}[Y_1] = \mathrm{E}[Y_2] = 3.5$. Thus, $\mathrm{E}[Y] = 3.5 + 3.5 = 7$.

Generalization: For any $n \geq 1$, the expectation of the sum of the face values of n fair dice $= 3.5 \times n$.