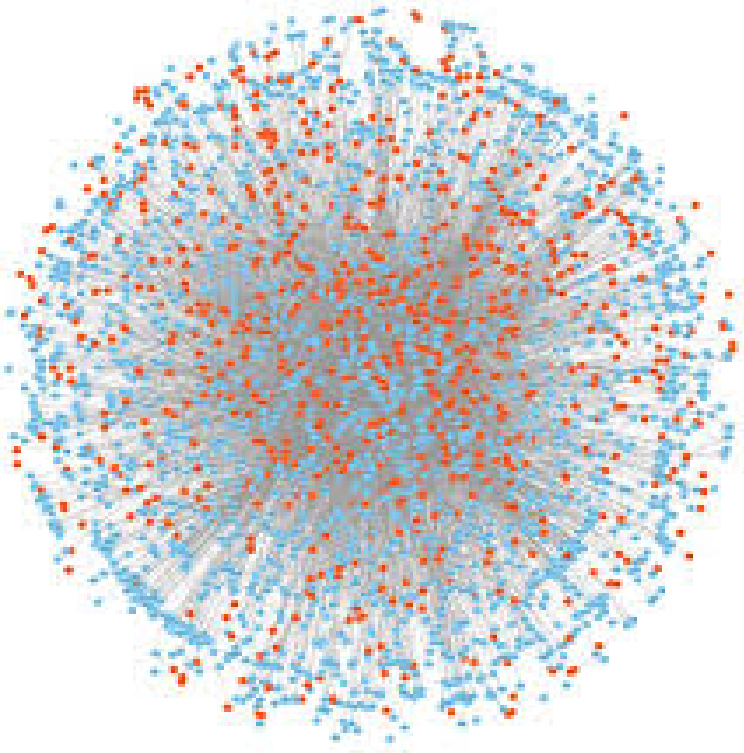
Shivan Ajay Iyer be22b048

WS Networks and E. coli Transcription

Assignment 2



PS1: Watts-Strogatz network vs Random Network

Methodology Overview

1. Network Generation

- Watts-Strogatz (WS):

  3 networks created with N=100, K=5 (10 total neighbors symmetrically), and β= p ∈ {0.3, 0.7, 1.0}

  - Uses ring lattice rewiring with probability β
  - Edge count preserved at 500 during rewiring

- Erdős-Rényi (ER):

  3 random networks were created with N=100 and exactly M=500 edges (same as WS networks)

2. Key Metrics Calculated

- Degree distribution: Histogram of node degrees normalised to probabilities
- Average clustering coefficient: Mean of local clustering coefficients
- Characteristic path length: Average shortest path between all connected node pairs

3. Implementation

- Built-in MATLAB functions for graph generation (*WattsStrogatz, graph*)
- Custom functions:
  - *clustering_coefficients* for node-level clustering
  - *ErdosRenyi* for ER graph generation with exact edge control

Key Findings for Immediate Reporting

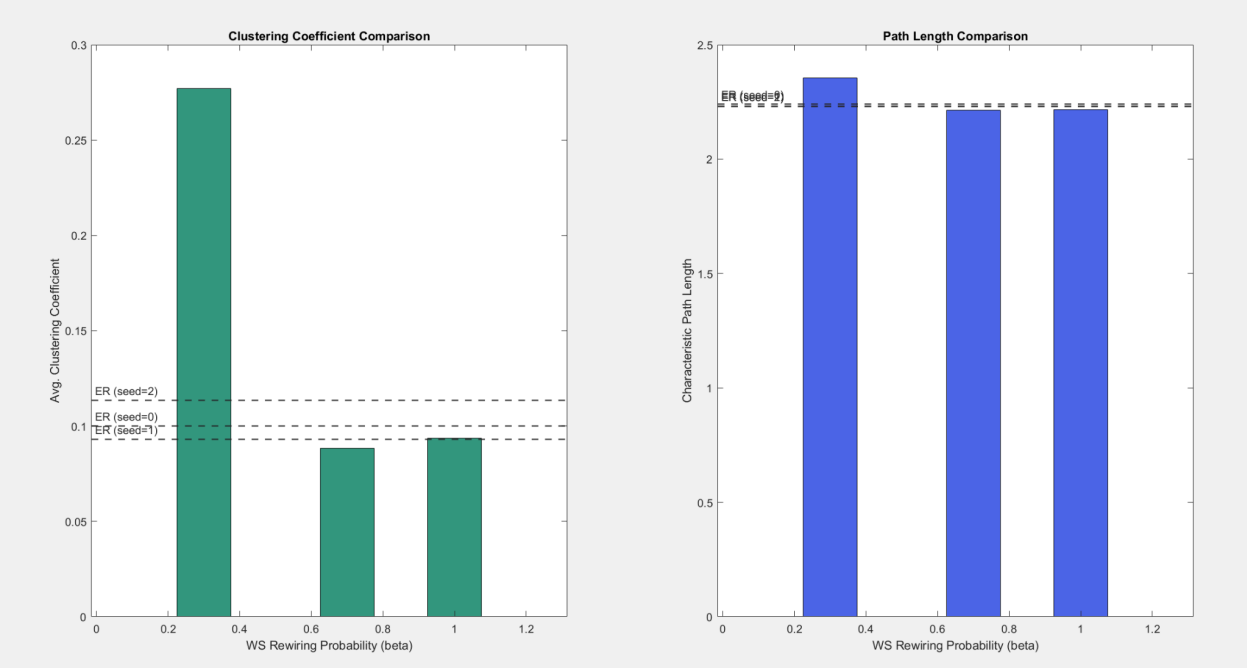
1. **Degree Distributions**
    - WS networks retain *near-regular structure* at β=0.3 but develop *Poisson-like distributions* as β→1
    - ER networks show classic *bell-shaped Poisson distribution*
2. **Clustering Coefficients**
    - WS networks maintain *higher clustering* than ER at low β (β=0.3 preserves community structure)
    - Clustering sharply decreases for WS as β→1, converging toward ER values
3. **Path Length Dynamics**
    - WS networks achieve *small-world properties* (short paths, high clustering) at mid β (0.3-0.7)
    - At β=1.0, WS path lengths match ER networks (random graph regime)
4. **Critical Transition**
    - At β=1.0, WS networks become *structurally indistinguishable* from ER graphs in:
        - Degree distribution shape
        - Global clustering magnitude
        - Average path length

Visualisation Strategy

The code generates two key comparison plots along with all degree distributions of the graphs:

1. **Clustering Coefficient vs β**
    - WS networks are shown as blue bars decreasing with β
    - ER networks as horizontal red baselines
2. **Path Length vs β**
    - WS path lengths (green bars) decline rapidly with β
    - ER values (red lines) mark the asymptotic lower-bound

These plots directly demonstrate the *small-world to random graph transition*. The comparison plot is shown below. All other plots of the degree distributions are attached in the zip file.



E. Coli Transcription Network Centrality Analysis

Network Construction

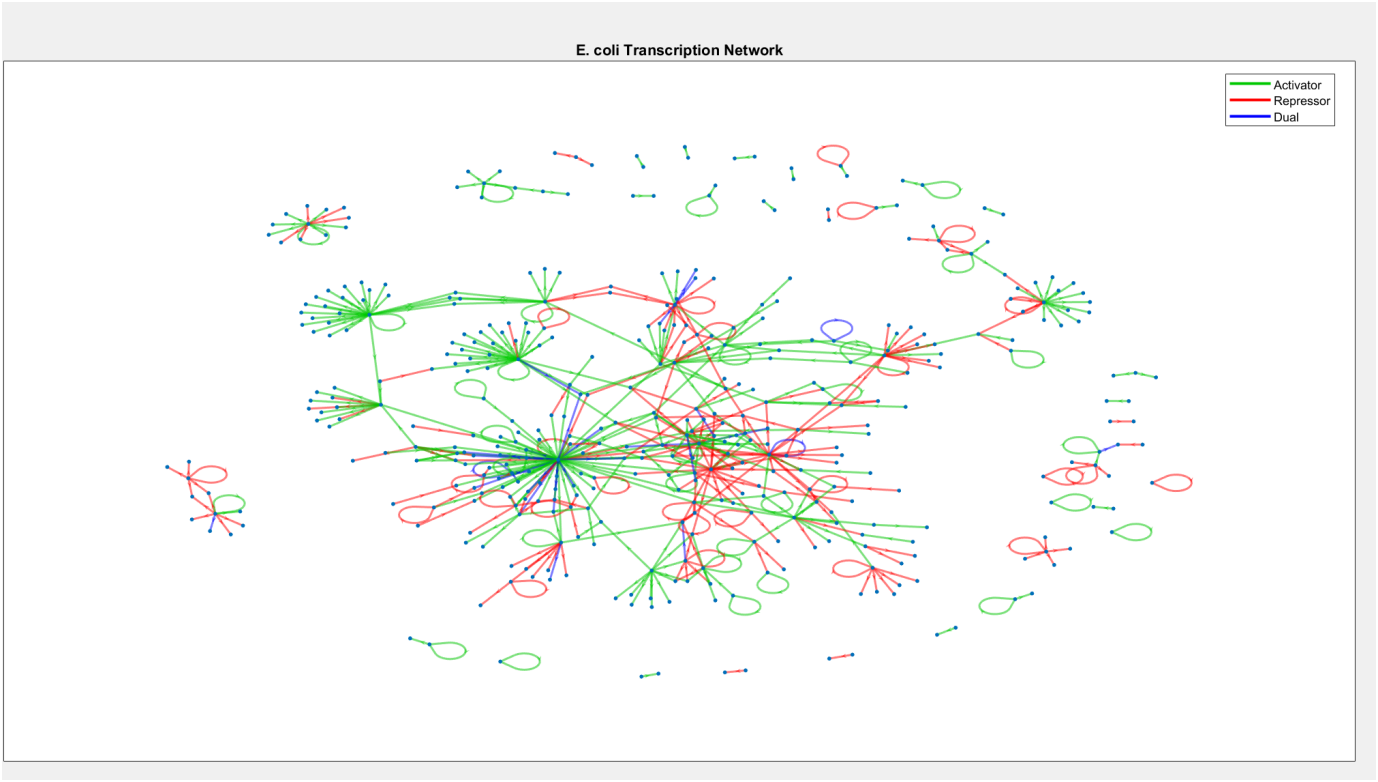
Data Processing

- Edge List Parsing:
  Directed graph built from *E_coli_transcription_network.txt* using MATLAB's *digraph*
    - Nodes: 2 types (Transcription Factors [TFs] and Operons)
    - Edges: Directed interactions labelled as activator, repressor, or dual

Visualisation Protocol

- Layout: Force-directed algorithm ('force' with gravity)
- Edge Coloring:
    - Activator: Green ([0, 0.8, 0])
    - Repressor: Red ([1])
    - Dual: Blue ([1])

Key Implementation Note
Closeness centrality is calculated using outcloseness paths originating from TFs), aligning with transcriptional regulation dynamics.



A. Centrality Metrics

Three measures were computed for all nodes:

1. Degree Centrality: Total edges (in + out)
2. Closeness Centrality: Reciprocal of average shortest path length from node to others
3. Betweenness Centrality: Fraction of shortest paths passing through a node

Filtering: Only TFs ranked (operons excluded).

B. Subnetwork Analysis

Activator Network: Activation Fraction Calculation:

$$\text{Fraction}_{\text{activation}} = \frac{\text{TF's activator edges}}{\text{TF's total outgoing edges}}$$

1. Top 5 Removed TFs:

   CRP (16.1%), rpoE_rseABC (7.5%), yhdG_fis (7.5%), fnr (4.8%), nlpD_rpoS (4.2%)

2. Edge Filtering: Retained *only activator edges* post-removal

Repressor/Dual Network: Repressor Fraction Calculation

$$\text{Fraction}_{\text{repression}} = \frac{\text{TF's repressor edges}}{\text{TF's total outgoing edges}}$$

1. Top 5 Removed TFs:

   purR (7.9%), arcA (7.5%), lexA_dinF (5.6%), fur (4.7%), himA (3.7%)

2. Edge Filtering: Retained *repressor + dual edges* post-removal

Results

Please find the detailed results in the zip folder in the file named ***ecoli_TF_command_terminal_data***. Below are some key insights and a snapshot of the results.

A. Network Visualisation

- The force-directed layout reveals CRP as a central hub.
- Regulation type distribution:

  `Activator (43%) | Repressor (34%) | Dual (23%)`

B. Centrality Rankings of the Original Network

Degree Centrality	Closeness Centrality	Betweenness Centrality
1. CRP (74)	1. CRP (0.00043481)	1. flhDC (49)
2. yhdG_fis (28)	2. rpoE_rseABC (0.00014753)	2. fliAZY (47)
3. rpoE_rseABC (26)	3. fnr (0.00014743)	3. rpoH (40)
4. fnr (24)	4. yhdG_fis (0.000146)	4. hns (22)
5. himA (23)	5. arcA (0.00011256)	5. ompR_envZ (18)

C. Activator Network Post Removal

1. Degree Distribution:
   - Range: 0–12 (vs. original 0–74)
   - Most frequent degree: 1 (41% nodes)
2. Top Degree Centrality:

   hns (10), rpoH (9), cpxAR (9)
3. Closeness Centrality Range: 0–0.00022883

D. Repressor/Dual Network Post Removal

1. Degree Distribution:
    - Bimodal peaks at 1 (44 nodes) and 3 (18 nodes)
2. Top Degree Centrality:

    fliAZY (12), fnr (12), marRAB (7)

3. Closeness Centrality Range: 0–0.00047959

Key Biological Insights

1. CRP as 'Master Regulator':
    - The highest degree (74) and closeness centrality (0.00043481) confirm the role of the master regulator.
    - The absence of betweenness centrality in the top 5 suggests a specialised regulatory role rather than structural bridging.
    - CRP Regulator Modulates Multidrug Resistance of Escherichia coli by Repressing the mdtEF Multidrug Efflux Genes. This paper confirms the conclusion drawn about the role of CRP from network analysis alone.
2. Regulatory Bottlenecks:
    - flhDC (betweenness=49) and fliAZY (47)

Network Resilience:

1) Activator removal reduces max degree by 84% (74→12).

    Biological Implication:
    Activator-rich networks rely on a few "master regulators" – their removal collapses coordinated gene expression.

2) Repressor removal increases betweenness spread (marRAB=7 shows emerging control points), i.e. control in the original network is more centralised (CRP/flhDC dominate). Still, post-repressor removal, there is decentralised control (marRAB gains influence).

   Biological Implication:
   Repression-dominated networks develop backup control points under stress.