INF6018 Project - Gender Equality, Education, and Economic Opportunities - A Global Network Analysis

Introduction

This project aims to analyze the relationships between gender equality, education, and economic opportunities/development across countries in Europe, Asia, Middle East, North Africa and North America. The study will focus on 61 countries, examining a specific gender equality metric and their correlation with education. The gender statistic data used is from the year 2022. The nodes are countries in which the gender statistics for Educational Attainment, Labor For Participation, (Un-)Employment Rates is available in the year 2022. The edges connect countries between their similiarity in the chosen gender equality metric. Additionally, data on each country is used to show the educational attainment after the threshold at least completed lower-secondary.

Research Questions:

1. Which groups of countries are similiar in their gender equality?
2. Which countries offer gender equality in their economic opportunities, which more and which less?
3. How do these clusters relate to education attainment?

Methodology

Node Attributes

```
'Labor': [
    'Ratio of female to male labor force participation rate (%) (modeled ILO estimate)',
    'Labor force participation rate, female (% of female population ages 15+) (modeled ILO estimate)',
    'Labor force participation rate, male (% of male population ages 15+) (modeled ILO estimate)'
],
'Employment': [
    'Employment to population ratio, 15+, female (%) (modeled ILO estimate)',
    'Employment to population ratio, 15+, male (%) (modeled ILO estimate)',
    'Unemployment, female (% of female labor force) (modeled ILO estimate)',
    'Unemployment, male (% of male labor force) (modeled ILO estimate)'
],
'Law': [
    'Women Business and the Law Index Score (scale 1-100)',
    'Women, Business and the Law: Pay Indicator Score (scale 1-100)',
    'Women, Business and the Law: Workplace Indicator Score (scale 1-100)',
    'Women, Business and the Law: Mobility Indicator Score (scale 1-100)'
],
'Education': [
    'Educational attainment, at least completed upper secondary, population 25+, female (%) (cumulative)',
    'Educational attainment, at least completed upper secondary, population 25+, male (%) (cumulative)',
    'Educational attainment, at least completed post-secondary, population 25+, female (%) (cumulative)',
    'Educational attainment, at least completed post-secondary, population 25+, male (%) (cumulative)',
    'Educational attainment, at least Bachelor\'s or equivalent, population 25+, female (%) (cumulative)',
    'Educational attainment, at least Bachelor\'s or equivalent, population 25+, male (%) (cumulative)'
],
'Economic': [
    'GDP per capita (constant 2010 US$)',
    'GNI per capita, Atlas method (current US$)',
    'GNI per capita, PPP (current international $)'
]
```

Similarity Calculation

- Develop a composite similarity metric using multiple attributes, which estimates the relationship between gender equality in economic opportunities for each country
- Compare countries based on this metric for the strength of this relationship

To create meaningful metrics, we'll use female-to-male ratios for most indicators to capture relative gender differences:

Economic Opportunities Gender Equality Index (EOGEI)

To create a metric for gender equality in economic opportunities using labor force participation and employment indicators, we can create a composite index:

Gender Parity in Labor Force Participation (GPLFP)

```
$$GPLFP = {\text{Female Labor Force Participation Rate} \over \text{Male Labor Force Participation Rate}} * 100$$
```

Gender Parity in Employment-to-Population Ratio (GPETP)

```
$$GPLFP = {\text{Female Employment-to-Population Ratio} \over \text{Male Employment-to-Population Ratio}} * 100$$
```

Gender Parity in Unemployment Labor Force Rate (GPULF)

```
$$GPLFP = {\text{Male Unemployment Rate} \over \text{Female Unemployment Rate}} * 100$$
```

Note: For GPULF, we use male/female to ensure that lower female unemployment results in a higher score.

Interpretation

- EOGEI = 100: Perfect gender parity
- EOGEI < 100: Disparity favoring males
- EOGEI > 100: Disparity favoring females

Gender parity is considered achieved if 97 ≤ EOGEI ≤ 103.

This composite metric provides a balanced view of gender equality in economic opportunities by considering participation, employment, and unemployment factors. It allows for a more nuanced understanding of gender disparities in the labor market, as it captures both the willingness to work (labor force participation) and the actual employment outcomes. Using both labor force participation and employment indicators captures different aspects of economic opportunities:

1. Labor force participation shows the proportion of the working-age population that engages in the labor market, either by working or looking for work.
2. Employment indicators provide information on the actual success in securing employment and the quality of labor market outcomes.

Gender Parity Index (GPI) for Education

```
$$EAGPI_{secondary} = {\text{Female Secondary Attainment (\%)} \over \text{Male Secondary Attainment (\%)}}$$
$$EAGPI_{postsec} = {\text{Female Postsecondary Attainment (\%)} \over \text{Male Postsecondary Attainment (\%)}}$$
$$EAGPI_{bachelor} = {\text{Female Bachelor's Attainment (\%)} \over \text{Male Bachelor's Attainment (\%)}}$$
```

Interpretation

- GPI = 1: Perfect gender parity
- GPI < 1: Disparity in favor of males
- GPI > 1: Disparity in favor of females

To check for gender equality, we use these thresholds:

- 0.97 ≤ GPI ≤ 1.03: Gender parity achieved
- GPI < 0.97: Disparity favoring males
- GPI > 1.03: Disparity favoring females

Overall Educational Attainment GPI To get a more comprehensive view of educational attainment across genders, we calculate an overall Educational Attainment GPI by averaging the three GPIs:

```
$$EAGPI_{overall} = {{EAGPI_{secondary}} + {EAGPI_{postsec}} + {EAGPI_{bachelor}} \over 3}$$
```

```
# Write edges
fout.write("*Edges\n")
for i in range(len(df_columns)):
    for j in range(i + 1, len(df_columns)):
        country1 = df_columns.index[i]
        country2 = df_columns.index[j]
        eogei1 = df_columns.iloc[i]
        eogei2 = df_columns.iloc[j]

        # Calculate the absolute difference
        diff = abs(eogei1 - eogei2)

        # Define the maximum allowed difference for connection
        max_diff = 7 # This can be adjusted based on your needs

        if diff <= max_diff:
            # Calculate weight: inverse of the difference, normalized
            weight = 1 - (diff / max_diff)
            fout.write(f"{nodes_dict[country1]} {nodes_dict[country2]} {weight:.4f}\n")
```

By using this approach countries with very similar EOGEI values will have edge weights closer to 1, representing a denser or stronger connection and instead of

discrete categories, this provides a continuous scale of relationships between countries. So, the weight directly represents how similar two countries are in terms of their EOGEI values. A weight of 0.9, for example, indicates a very strong similarity, while a weight of 0.1 suggests the countries are at the edge of being considered connected at all.

Implications

- Countries above 107 can be easily identified as leaders in gender equality, while those below 93 are lagging behind.
- Countries within the thresholds are of particular interest, as they represent transitional states in gender equality progress.
- Identify regions or groups of countries with similar gender equality levels more easily.

Results

Network Parameters of Node Centrality

Excerpt from RStudio Console:

```
> top_nodes_degree
  Bulgaria Luxembourg  Australia    Denmark      Malta
        30         30         28         28         28
> top_nodes_betw
               Czechia Bosnia and Herzegovina               Slovenia                  India
              344.3333               277.0000               220.3780              196.0000
               Croatia
              188.7989
> top_nodes_clo
           Lao PDR            Moldova West Bank and Gaza            Bahrain Iran, Islamic Rep.
         3.2226877          3.2226877          0.7789375          0.5393161          0.4395604
> top_nodes_eigen
       Sweden       Hungary United States       Ireland      Bulgaria
    1.0000000     0.9999475     0.9956765     0.9919266     0.9908859
```

Degree Centrality

The top countries by degree centrality are:

- Bulgaria
- Luxembourg
- Australia
- Denmark
- Malta These countries have the highest number of connections/edges to other countries in the network, indicating that they have similar gender equality metrics to many other nations. This suggests that these countries may represent "average" or "typical" cases in terms of gender equality measures across countries.

Betweenness Centrality

The top countries by betweenness centrality are:

- Czechia
- Bosnia and Herzegovina
- Slovenia
- India
- Croatia These countries act as important bridges or connectors in the network. They likely represent transitional states between different clusters of countries with varying levels of gender equality. Their position suggests they may be important in understanding how gender equality metrics transition between different groups of countries.

Closeness Centrality

The top countries by closeness centrality are:

- Lao PDR
- Moldova
- West Bank and Gaza
- Bahrain
- Iran, Islamic Rep. These countries are, on average, closest to all other countries in the network in terms of their gender equality metrics. This suggests that they may represent either "median" or "outlier" cases in the study, potentially offering insights into the typical gender equality situations in their diverse regions.

Eigenvector Centrality

The top countries by eigenvector centrality are:

- Sweden
- Hungary
- United States
- Ireland
- Bulgaria These countries are not only well-connected themselves but are also connected to other highly connected countries. This suggests that they are central to clusters of nations with similar gender equality profiles, potentially representing leaders within their respective regions or development categories.

Interpretation and Implications

- The appearance of countries from various regions (Europe, Asia, Middle East, North America) across different centrality measures highlights the complex and diverse nature of gender equality issues globally.
- Some unexpected countries appear as central in different measures (e.g., Lao PDR in closeness centrality), which may indicate unique or transitional positions in gender equality metrics for further investigation.
- Countries with high betweenness centrality (like Czechia and Bosnia and Herzegovina) might be particularly interesting for policymakers as they could represent transitional states in gender equality progress.
- The presence of both highly developed (e.g., United States, Sweden) and developing countries (e.g., India, Lao PDR) in various centrality measures suggests that the relationship between economic opportunities and gender equality is more complex and not necessarily linear.

Centralization Scores

Excerpt from RStudio Console:

```
> network_centr_degree
[1] 0.236612
> network_centr_betw
[1] 0.113491
> network_centr_clo
[1] NaN
> network_centr_eigen
[1] 0.5547316
```

Degree Centralization: 0.236612

- A decentralized structure in terms of connections (no domination from a country/countries in the network connection-wise)
- Similarity in the gender equality metric is more or less evenly distributed across countries The relatively low score indicates that gender equality is relatively diverse across the countries.

Betweenness Centralization: 0.113491

- Multiple countries serve as bridges between different parts of the network.
- The network doesn't rely on a few key countries to connect different clusters. The low score suggests that there are multiple pathways of similarity or transition between different groups of countries. Various countries play intermediary roles in connecting different levels or types of gender equality profiles.

Closeness Centralization: NaN

- A disconnected component in the network (there are vertices that cannot reach all other vertices in the network) It could indicate a methodological issue or maybe a unique network structure.

Identify where the NaN value is coming from

```
> nan_vertices <- which(is.nan(closeness_cent))
> nan_vertices
Afghanistan
          1
```

Eigenvector Centralization: 0.5547316

- There is a hierarchy in the network.
- Some countries are more central (connected to other well-connected countries) than others. This suggests that there are some countries that stand out as leaders or central figures in gender equality metrics. These countries are not only similar to many others but are also similar to other influential countries. This could represent a core group of nations that are at the forefront of gender equality measures or economic opportunities related to gender issues.

Interpretation and Implications

- The low degree and betweenness centralization values suggest a diverse landscape of gender equality profiles across countries, without dominant central players.
- There are likely multiple countries influencing or bridging different aspects of gender equality and development, rather than a few dominant influencers.
- The higher eigenvector centralization value suggests a possible core-periphery structure, where some countries form a central, interconnected group in terms of gender equality measures.
- The closeness centralization score warrants further investigation. It might indicate a unique network structure or potentially a methodological consideration in how similarities between countries are measured.

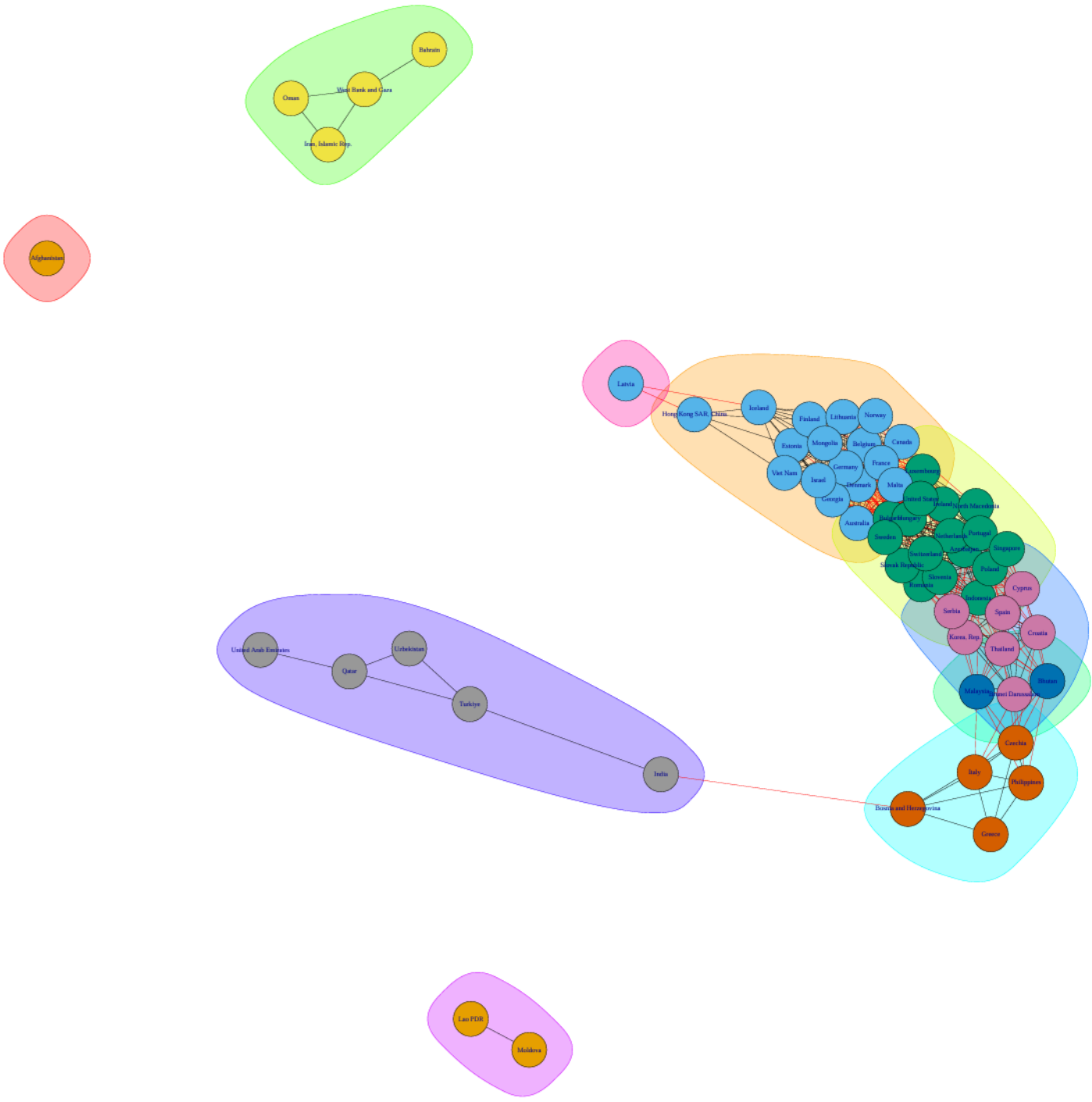
Community Characteristics

Community by Edge Betweenness

```
IGRAPH clustering edge betweenness, groups: 10, mod: 0.41
+ groups:
  $`1`
  [1] "Afghanistan"

  $`2`
   [1] "Australia"           "Belgium"             "Canada"              "Denmark"
   [5] "Estonia"             "Finland"             "France"              "Georgia"
   [9] "Germany"             "Hong Kong SAR, China" "Iceland"            "Israel"
  [13] "Lithuania"           "Malta"               "Mongolia"            "Norway"
  [17] "Viet Nam"

  + ... omitted several groups/vertices
 [1]  1  2  3  4  2  5  6  7  3  2  7  7  6  2  2  2  2  2  2  6  2  3  2  8  3  4  3  2  6  7  9 10  2
[34]  3  5  2  9  2  3  3  2  4  6  3  3  8  3  7  3  3  3  7  3  3  7  8  8  3  8  2  4
```



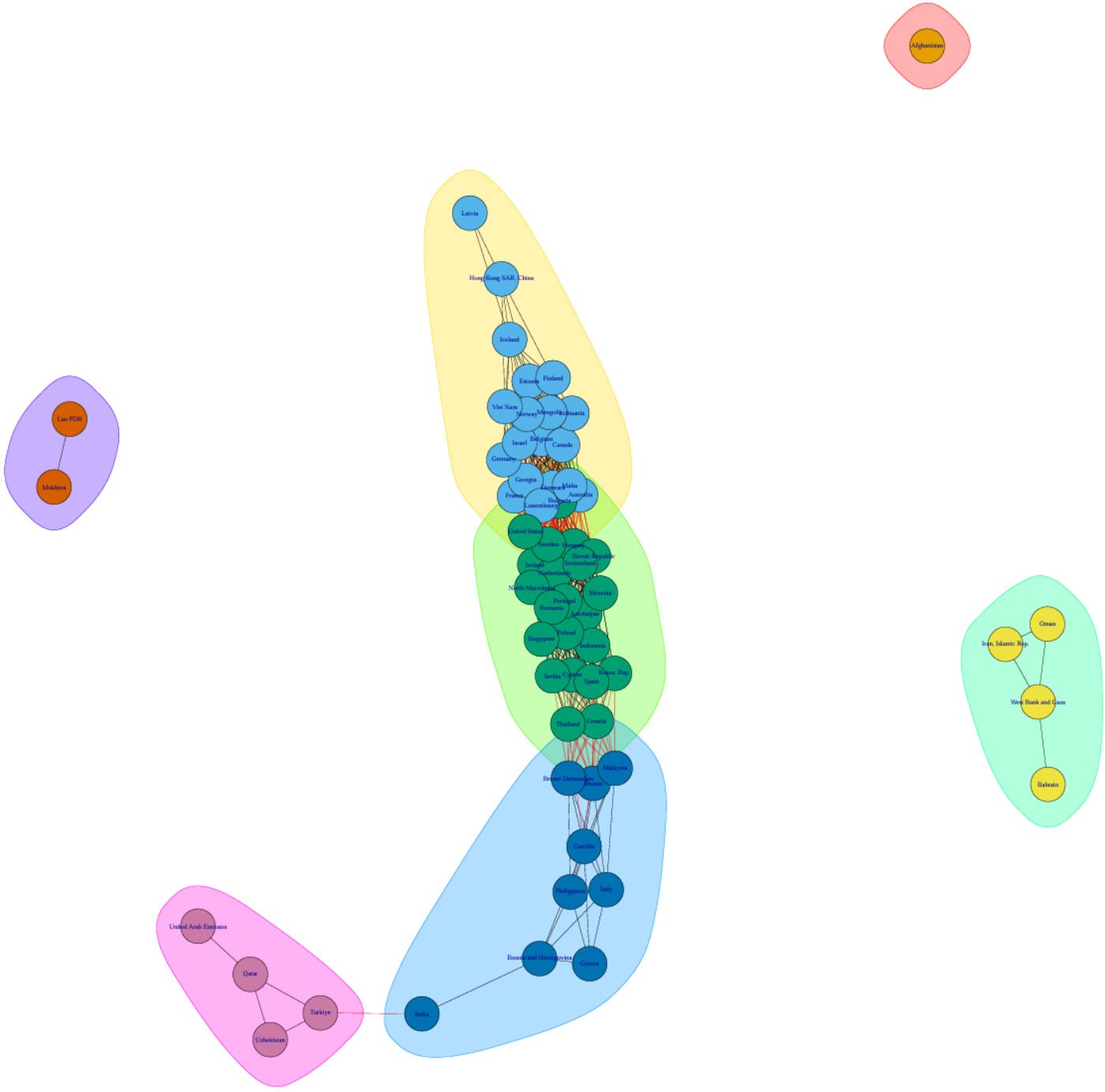
There's a noticeable tendency for countries from similar geographic regions to cluster together, suggesting regional similarities in gender equality progress. The largest community seems to consist of highly developed nations, indicating a correlation between economic opportunities and the gender equality metric. Both blue communities' composition suggests that cultural/religious factors may play a role in shaping gender equality profiles in some regions.

Community by Label Prop

```
IGRAPH clustering label propagation, groups: 7, mod: 0.42
+ groups:
  $`1`
  [1] "Afghanistan"

  $`2`
   [1] "Australia"           "Belgium"             "Canada"              "Denmark"
   [5] "Estonia"             "Finland"             "France"              "Georgia"
   [9] "Germany"             "Hong Kong SAR, China" "Iceland"            "Israel"
  [13] "Latvia"              "Lithuania"           "Luxembourg"          "Malta"
  [17] "Mongolia"            "Norway"              "Viet Nam"

  + ... omitted several groups/vertices
 [1] 1 2 3 4 2 5 5 5 3 2 3 3 5 2 2 2 2 2 2 5 2 3 2 5 3 4 3 2 5 3 6 2 2 2 5 2 6 2 3 3 2 4 5 3 3 7 3 3 3 3
[51] 3 3 3 3 3 7 7 3 7 2 4
```



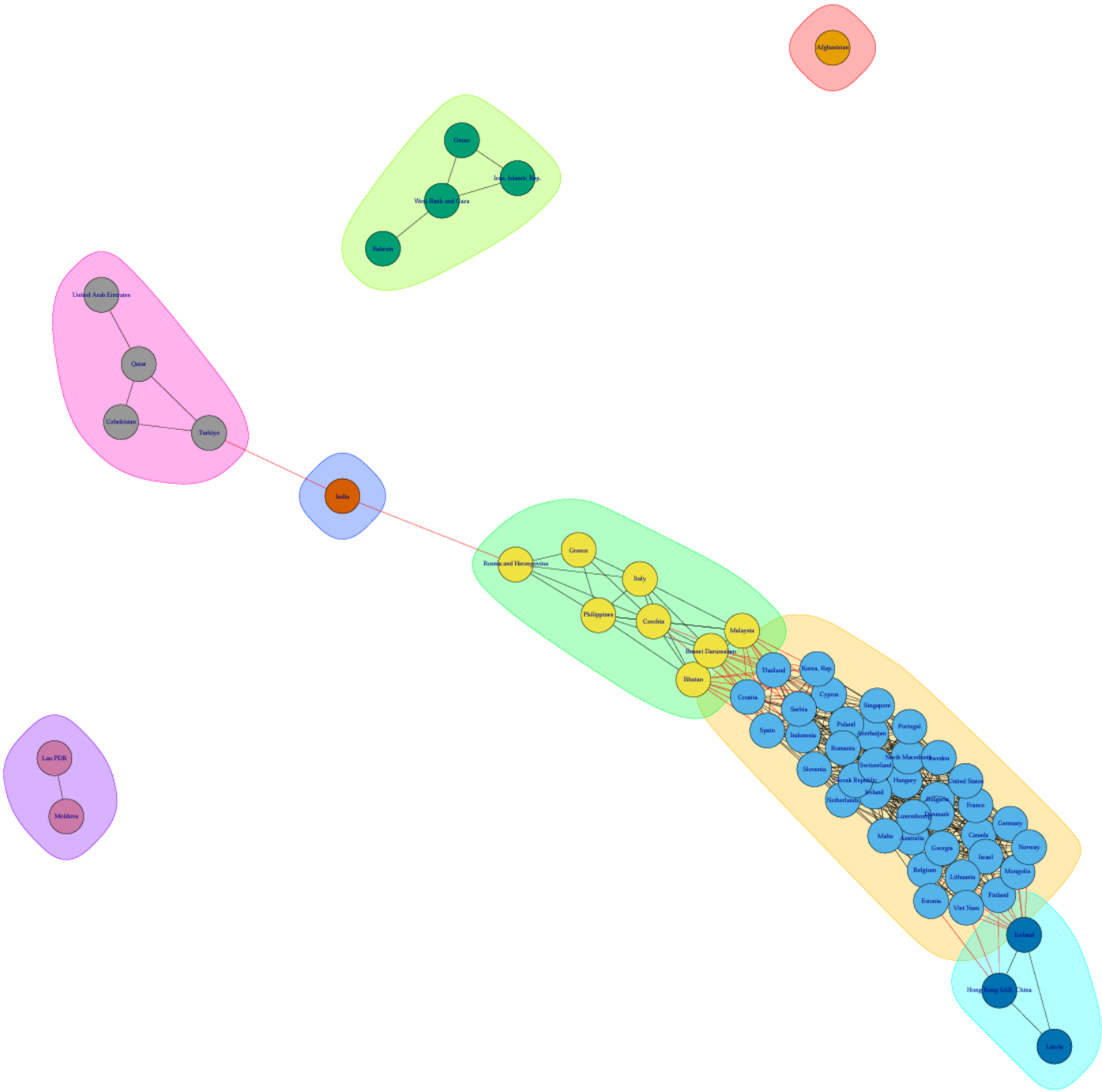
Compared to the edge betweenness method, this algorithm produced fewer, but larger communities, suggesting more overarching similarities in gender equality patterns. The green and yellow communities appear to group highly developed or fast developing nations, indicating a strong correlation between economic opportunities and gender equality metric, while the blue community seems to represent countries at different stages of the gender equality progress.

Community by Leiden

```
 "resolution"
       25%
0.03566667
IGRAPH clustering leiden, groups: 8, mod: NA
+ groups:
  $`1`
  [1] "Afghanistan"

  $`2`
   [1] "Australia"       "Azerbaijan"      "Belgium"         "Bulgaria"        "Canada"
   [6] "Croatia"         "Cyprus"          "Denmark"         "Estonia"         "Finland"
  [11] "France"          "Georgia"         "Germany"         "Hungary"         "Indonesia"
  [16] "Ireland"         "Israel"          "Korea, Rep."     "Lithuania"       "Luxembourg"
  [21] "Malta"           "Mongolia"        "Netherlands"     "North Macedonia" "Norway"
  [26] "Poland"          "Portugal"        "Romania"         "Serbia"          "Singapore"
  + ... omitted several groups/vertices
 [1] 1 2 2 3 2 4 4 4 2 2 2 2 4 2 2 2 2 2 2 4 5 2 5 6 2 3 2 2 4 2 7 5 2 2 4 2 7 2 2 2 2 3 4 2 2 8 2 2 2 2
[51] 2 2 2 2 2 8 8 2 8 2 3
```

The Leiden algorithm has produced more nuanced communities compared to the previous methods with finer distinctions in gender equality patterns. All the green and purple communities seem to represent different levels of advanced gender equality, possibly distinguishing between established and cutting-edge approaches. There's also a noticeable regional clustering, particularly in the yellow and blue communities, which shows strong regional factors in gender equality progress.



Node Attributes and Clustering

We clustered countries using the overall EAGPI in Orange using k-Means. We obtained a list of countries assigned to clusters. Countries in clusters are different based on EAGPI value.

```
print(names(education))
```

```
[1] "Country Name"  "Cluster"       "Silhouette"    "EAGPI_overall"
```

```
edu_clusters<-data.frame(education$CountryName, education$Cluster, education$EAGPI_overall)
community_ids<-data.frame(ldc$names, ldc$membership)

merged <- merge(x = community_ids, y = edu_clusters, by.x = "ldc.names", by.y = "education.CountryName")
merged
```

	ldc.names	ldc.membership	education.Cluster	education.EAGPI_overall
1	Afghanistan	1	C3	0.31
2	Australia	2	C2	1.12
3	Azerbaijan	2	C4	0.93
4	Bahrain	3	C2	1.26
5	Belgium	2	C4	1.05
6	Bhutan	4	C1	0.9
7	Bosnia and Herzegovina	4	C4	0.94
8	Brunei Darussalam	4	C2	1.14
9	Bulgaria	2	C2	1.23
10	Canada	2	C4	1.05
11	Croatia	2	C2	1.11
12	Cyprus	2	C4	1.03
13	Czechia	4	C4	1
14	Denmark	2	C2	1.23
15	Estonia	2	C2	1.32
16	Finland	2	C2	1.15
17	France	2	C4	1.01
18	Georgia	2	C4	1.07
19	Germany	2	C1	0.87
20	Greece	4	C4	0.97
21	Hungary	2	C4	1.07
22	Iceland	5	C2	1.26
23	India	6	C1	0.66
24	Indonesia	2	C4	1.01
25	Ireland	2	C4	1.07
26	Israel	2	C2	1.13
27	Italy	4	C2	1.15
28	Lao PDR	7	C1	0.69
29	Latvia	5	C2	1.26
30	Lithuania	2	C2	1.17
31	Luxembourg	2	C4	0.94
32	Malaysia	4	C4	0.99
33	Malta	2	C4	0.99
34	Moldova	7	C2	1.14
35	Mongolia	2	C2	1.19
36	Netherlands	2	C4	0.96
37	North Macedonia	2	C4	1
38	Norway	2	C2	1.21
39	Oman	3	C5	1.54
40	Philippines	4	C2	1.19
41	Poland	2	C2	1.22
42	Portugal	2	C2	1.23
43	Qatar	8	C5	1.79
44	Romania	2	C4	1.01
45	Serbia	2	C4	0.99
46	Singapore	2	C4	1.04
47	Slovak Republic	2	C2	1.15
48	Slovenia	2	C2	1.2
49	Spain	2	C2	1.11
50	Sweden	2	C2	1.2
51	Switzerland	2	C1	0.83
52	Thailand	2	C2	1.11
53	Turkiye	8	C1	0.81
54	United Arab Emirates	8	C4	1
55	United States	2	C4	1.06
56	Uzbekistan	8	C1	0.86
57	Viet Nam	2	C1	0.81
58	West Bank and Gaza	3	C2	1.12

```
print("How many clusters")
[1] "How many clusters"
print(unique(merged$education.Cluster) )
[1] "C3" "C2" "C4" "C1" "C5"

print("How many communities")
[1] "How many communities"
print(unique(merged$ldc.membership) )
[1] 1 2 3 4 5 6 7 8
```

Not possible to clearly associate education cluster with the community.

```
table(merged$education.Cluster,merged$ldc.membership)

     1  2  3  4  5  6  7  8
  C1  0  3  0  1  0  1  1  2
  C2  0 17  2  3  2  0  1  0
  C3  1  0  0  0  0  0  0  0
  C4  0 17  0  4  0  0  0  1
  C5  0  0  1  0  0  0  0  1
```

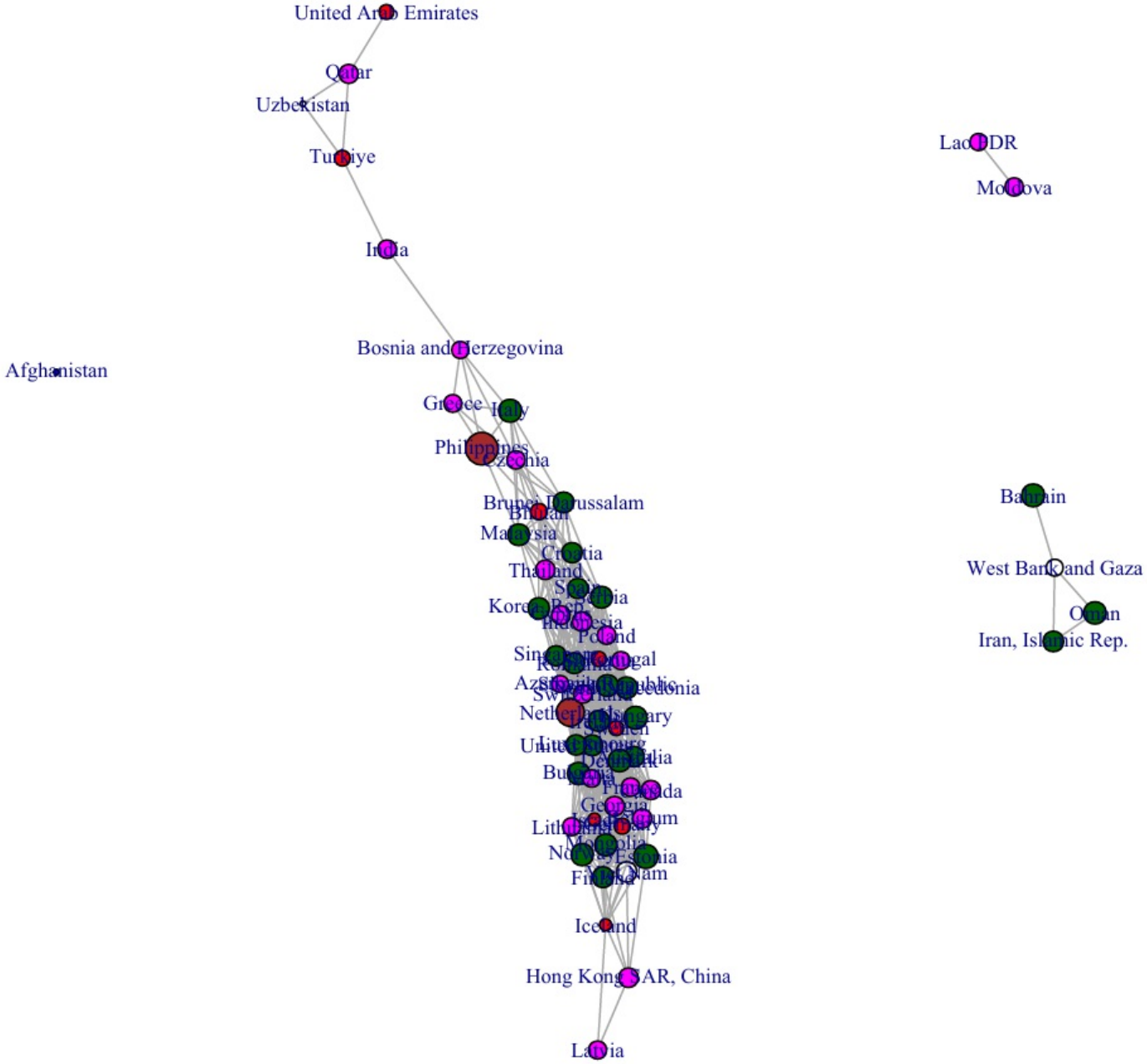
Map education information on the network

```
# Map color onto the vertices
clusters<-merged$education.Cluster
names(clusters) <- merged$ldc.names
#print(clusters)
cluster_colors<-c("red","darkgreen","blue","magenta","brown","cyan")
names(cluster_colors)<-c("C1","C2","C3","C4","C5",NA)

# Map size onto the vertices
EAGPI<-as.double(merged$education.EAGPI_overall)
names(EAGPI)<-merged$ldc.names

V(network)$label.cex <- 0.7

plot(network,
     layout = layout.graphopt,
     vertex.color = cluster_colors[clusters[V(network)]],
     vertex.size = rep(EAGPI, length.out = vcount(network)) * 3.5)
```



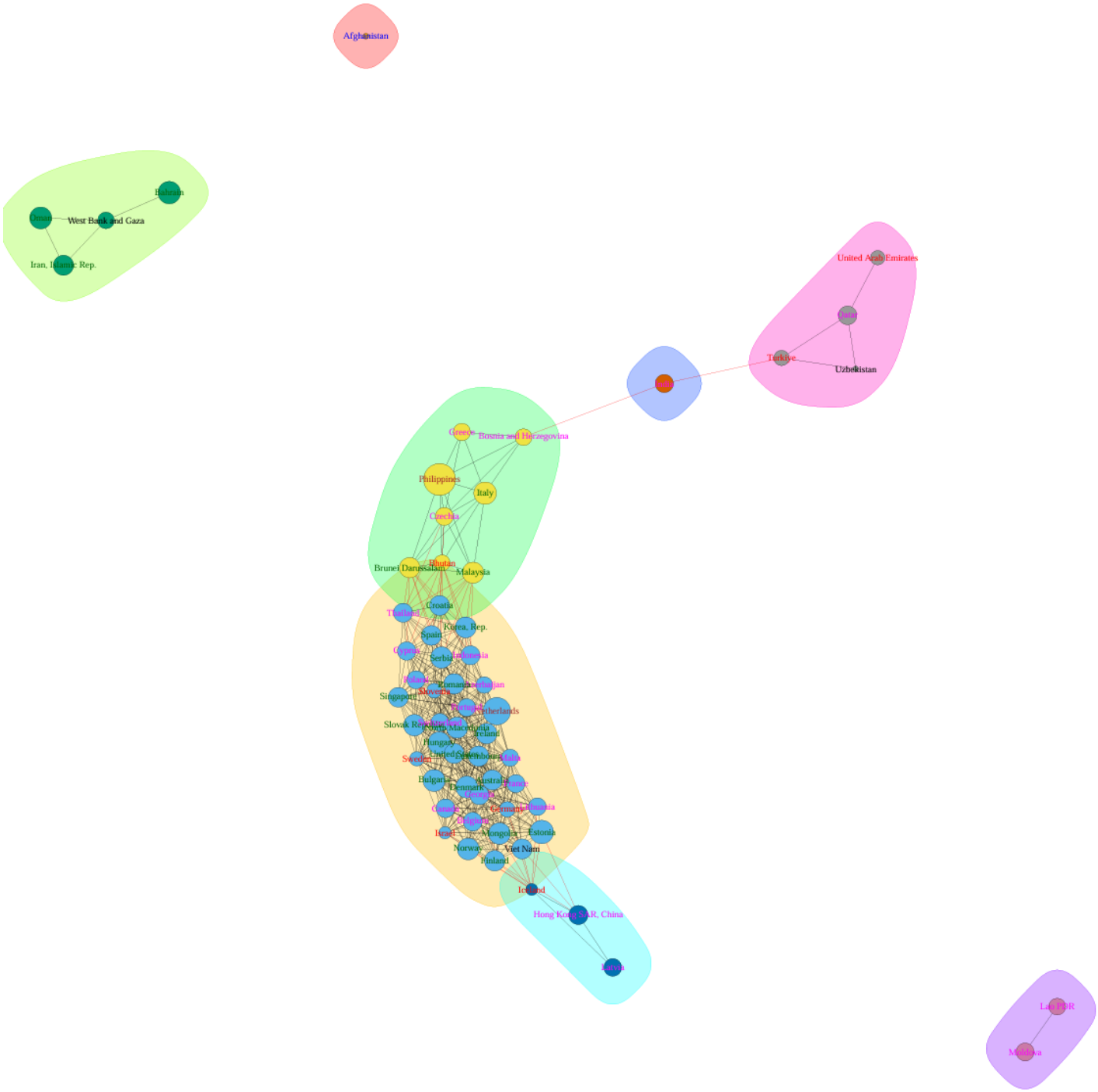
The size of each node represents the Educational Attainment Gender Parity Index (EAGPI) value, while the color indicates the educational cluster to which each country belongs.

Larger nodes (higher EAGPI values) tend to be more centrally located in the network. This could suggest a positive correlation between educational gender parity and overall gender equality in economic opportunities. Some of the largest nodes show exceptionally high educational gender parity, especially those in the Gulf states. Smaller nodes are often found on the periphery of the network, which suggests that countries with lower educational gender parity may also have less similarity in overall gender equality metrics with other nations.

The network shows a mix of colors distributed throughout, which might indicate that countries with similar gender equality profiles in economic opportunities may have diverse levels of educational gender parity. There are definitly visible concentrations of certain colors, particularly green and pink, which likely represent the most common educational attainment patterns among the countries chosen. The presence of multiple colors within close proximity in the network suggests that educational gender parity does not always directly correspond to similarities in economic gender equality.

Add the community information, and education cluster information

```
plot(ldc,
     network,
     layout = layout.graphopt,
     vertex.label.color = cluster_colors[clusters[V(network)]],
     vertex.size = rep(EAGPI, length.out = vcount(network)) * 3.5)
```



The communities (represented by different colors in the network) often contain countries from various educational attainment clusters (indicated by node colors and sizes). This suggests that countries with similar gender equality profiles in economic opportunities may have diverse levels of educational gender parity. Countries with higher EAGPI values tend to occupy more central positions in the network, which could again imply a positive correlation between educational gender parity and overall gender equality in economic opportunities. There are also visible regional clusters in the network, particularly for European and North American countries (often grouped together) and Middle Eastern/North African nations (forming their own cluster). However, these regional groupings are not strictly aligned with educational attainment clusters, showing that regions or shared cultural backgrounds do not necessarily translate to similar educational gender parity levels. Afghanistan appears as an isolated node, suggesting significant differences in both economic and educational gender equality compared to other countries. Some countries, like Qatar and Oman, show high EAGPI values but are positioned differently in the network, indicating unique combinations of educational and economic gender equality. A large cluster of developed nations (mostly European countries, the US, and Canada) shows a mix of high EAGPI values (large nodes) and varying educational cluster assignments (different node colors), underlining that even among economically advanced countries, there are nuances in educational gender parity. Countries from Eastern Europe and Central Asia often form a distinct community but show varied educational attainment levels.

Conclusions

1. Which groups of countries are similar in their gender equality?

   A large group of developed nations, primarily from Western Europe and North America, consistently clustered together across different community detection methods. This suggests a strong similarity in gender equality measures among economically advanced countries. Eastern European and Central Asian countries often formed a separate community, indicating a distinct gender equality landscape in transitional economies. Middle Eastern and North African nations frequently grouped together, pointing to shared challenges or cultural factors influencing gender equality in this region.Some unexpected groupings emerged, such as the cluster containing Lao PDR, Moldova, and West Bank and Gaza, which may represent countries with unique gender equality profiles that warrant further investigation.

2. Which countries offer gender equality in their economic opportunities, which more and which less?

Countries like Sweden, Hungary, the United States, Ireland, and Bulgaria showed high eigenvector centrality, suggesting they are central to clusters of nations with similar, likely advanced, gender equality profiles. The presence of countries like Lao PDR, Moldova, and West Bank and Gaza in the top closeness centrality measures indicates that these nations may represent either median cases or interesting outliers in gender equality metrics. Afghanistan consistently appeared as an outlier, often forming its own single-country community, indicating significant challenges in gender equality.

3. How do these clusters relate to education attainment?

   There is no clear one-to-one correspondence between the communities identified based on economic gender equality metrics and the clusters formed based on the Educational Attainment Gender Parity Index (EAGPI). However, the visualization of EAGPI data on the network graph shows that countries with higher EAGPI values (represented by larger node sizes) tend to be more centrally located in the network, suggesting a positive relationship between educational gender parity and overall gender equality in economic opportunities. The diversity of EAGPI values within communities indicates that countries with similar economic gender equality profiles may still have varying levels of educational gender parity.

References

- Igraph reference. https://cran.r-project.org/web/packages/igraph/vignettes/igraph.html
- World Bank Databank World Development Indicators: https://databank.worldbank.org/source/gender-statistics
- CEIC Data:
  - https://www.ceicdata.com/en/thailand/education-statistics/th-educational-attainment-at-least-completed-postsecondary-population-25-years-female--cumulative
  - https://www.ceicdata.com/en/thailand/education-statistics/th-educational-attainment-at-least-completed-postsecondary-population-25-years-male--cumulative
  - https://www.ceicdata.com/en/uzbekistan/education-statistics/uz-educational-attainment-at-least-completed-postsecondary-population-25-years--cumulative-female
  - https://www.ceicdata.com/en/uzbekistan/education-statistics/uz-educational-attainment-at-least-completed-postsecondary-population-25-years--cumulative-male
- https://toreopsahl.com/2010/03/20/closeness-centrality-in-networks-with-disconnected-components/