SSNA 1-5 Deliverable: Co-authorship Social Network Analysis

Alice (Shuhan Ai)

October 12, 2022

Dataset

Topic: Student Enrollment in Community College

Description of Dataset

I exported my data from Scopus. Began in basic search with Article tile, Abstract, and Keyword, my research topic is "Student Enrollment in Community College". The documents were limited to all open access articles in social science field, including 5 years research from 2018 to 2022.

Query String

TITLE-ABS- KEY (student AND enrollment AND in AND community AND college) AND (LIMIT-TO (OA , "all")) AND (LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018)) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (SUBJAREA , "SOCI"))

Deliverable 1

Deliverable 1 Q1: Describe the network, how many authors are represented, how many connections are in this network?

```
#load packages
library(splitstackshape)
library(igraph)

#read table
setwd("~/Data/SSNA")
a<-read.csv("scopus1.csv")

#Keeping only column with author relationships
authors<-as.data.frame(a[,1])
head(authors)
```

```
##                                          a[, 1]
## 1                          Liu V.Y.T., Fay M.P.
```

```
## 2                    Zhao K., park t., Mokhe C.G., Hu S.
## 3                Gonzales S.M., Baier S.T., Brammer E.C.
## 4 Naor M., Pinto G.D., Davidov P., Benaroch J., Kivel A.
## 5                    Grosz M., Kurlaender M., Stevens A.
## 6                                        Liu V., Xu D.
```

```
colnames(authors)<-"AU"

#As can be seen in the file the separator of interest is :
a1<-cSplit(authors, splitCols = "AU", sep = ",", direction = "wide", drop = FALSE)
#Here we just drop the original first column
a1<-a1[,-1]
class(a1)
```

```
## [1] "data.table" "data.frame"
```

```
#read it as a matrix
mat <- as.matrix(a1)

mat<-tolower(mat)

#dimision result 56, 7
dim(mat)
```

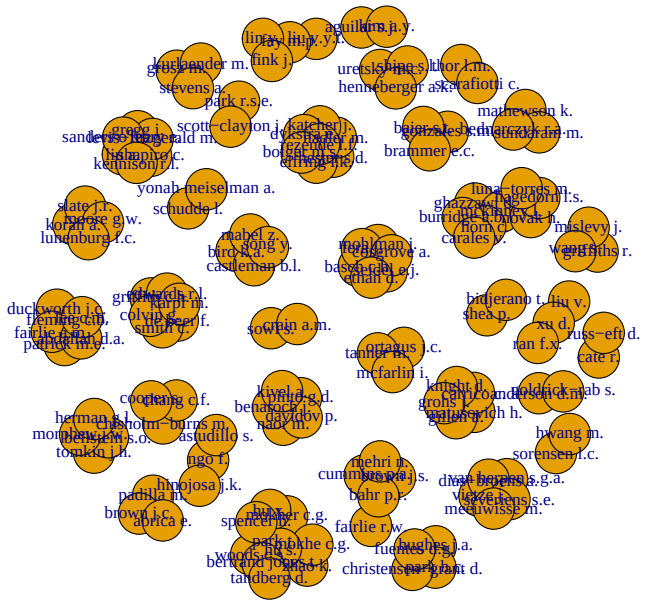
```
## [1] 56  7
```

```
# create edgelist _ all combinations except first author with itself
a1<-mat
edgelist1<-matrix(NA, 1, 2)#empty matrix two columns
for (i in 1:(ncol(a1)-1)) {
  edgelist11 <- cbind(a1[, i], c(a1[, -c(1:i)]))
  edgelist1 <- rbind(edgelist1,edgelist11)
  edgelist1<-edgelist1[!is.na(edgelist1[,2]),]
  edgelist1<-edgelist1[edgelist1[,2]!="",]
}
dim(edgelist1)
```

```
## [1] 231   2
```

The visualization of co-authorship network in R.

```
#plot author relationship
plot(graph.edgelist(edgelist1))
```



```
g<- graph.edgelist(edgelist1, directed = FALSE)
g.c <- simplify(g)
```

This is an one-mode unweighted network. There is total 137 authors are in the network, total 231 connections in the network, shown in the g. After condensing the relationship, 223 connections remain in g.c.

Deliverable 1 Q2: Who are the top five most prolific authors? What are they publishing about?

```
#Add publication as first column
a1 <- cbind(a$Source.title, a1)

mat <- as.matrix(a1)

edgelist_two_mode <- cbind(mat[, 1], c(mat[, -1]))
edgelist_two_mode <- edgelist_two_mode[!is.na(edgelist_two_mode[,2]), ]
g2<- graph.edgelist(edgelist_two_mode[, 2:1], directed = FALSE)  #this is from author to publication

#sort shows how many individual authors
V(g2)$type <- V(g2)$name %in% edgelist_two_mode[ ,1]
# degree(g2)[V(g2)$type==F]
# sort(degree(g2)[V(g2)$type==F], decreasing =T)
```

```
sort1 <-sort(degree(g2)[V(g2)$type==F], decreasing =T)
head_print(sort1, 10)
```

```
##              park t.                hu s.            bahr p.r.
##                    4                    4                    3
##               ngo f.               xu d.          mckinney l.
##                    2                    2                    2
##           woods c.s.            fay m.p.    bertrand jones t.
##                    2                    2                    2
##           liu v.y.t.             zhao k.        gonzales s.m.
##                    1                    1                    1
##              naor m.            grosz m.               liu v.
##                    1                    1                    1
```

There are two authors (Park T. and Hu S.) published 4 articles in this field, one author (Bahr P. R.) published 3 articles, and 6 authors (Ngo F.; Xu D.; Mckinney L.; Woods C. S.; Fay M. P.; Bertrand Jones T.) publish 2 articles each.

Park T. and Hu S. are coauthors of these four articles. By employing student-level data from two first-time-in-college cohorts, Zhao et al. (2022) analyze the impact of Senate Bill 1720 (SB 1720) on non-exempt students, indicating that non-exempt students benefit from the policy, with a significantly higher percentages of students enrolling in and completing college-level and advanced English and math courses after the reform. This research echoes the previous cohort-by-cohort comparative analysis, suggesting that SB 1720 may help to accelerate student success in community college, especially for Blank and Hispanic students Park-Gaghan et al. (2020). Furthermore, Woods et al. (2019) provide evident at the level of preparation was related to course enrollment and success. This research echoes the previous research on math course enrollment patterns for underprepared first-time-in-college (FTIC) students, indicating that students who received more developmental support and be more prepared were more likely to succeed in class Park et al. (2018).

Bahr P. R.'s research focuses on the labor market for community college. Leveraging data from California Employment Development Department (EDD) and Unemployment Insurance (UI) system, he found that strong, positive returns to completed credits in career and technical education (CTE) fields that are closely linked to employment sectors that are not credential-intensive, which could help colleges to target efforts to grow postsecondary completion opportunities for students through short-term certificates program Bahr (2019). And he further provides evidence that computer skills have short- or medium-run effects on earnings Fairlie and Bahr (2018). Besides, drawing on IPEDS data, he also explores how age would affect student enrollment patterns, showing that older learners tend to be more heterogeneous in attainment goals in community college Cummins et al. (2019).

The difference-in-difference method was used in Ngo F.'s research to provide evidence for undocumented community college student course enrollment after the California Dream act, showing that while state financial aid may have broadened postsecondary possibilities, access to aid did not increase undocumented students' credit loads to the level of their peers Ngo and Hinojosa (2022). He also investigates the impact of financial aid for undocumented community college students, suggesting undocumented students receiving aid achieved at similar levels as U.S. citizen peers receiving aid and better than their undocumented peers not receiving aid Ngo and Astudillo (2019).

Xu D. is another a prolific author in community college enrollment and student learning outcomes field. Using administrative data from a large state community college system and a two-way fixed effects model, the result indicates that greater exposure to dual enrollment (DE) peers during a student's initial semester in college reduces next-term college persistence V. Liu and Xu (2022). She also identifies the course enrollment and academic outcomes for noncredit students in community college, supporting that noncredit students tend to be adult learns and are typically from a lower socioeconomic background than credit students at community colleges Xu and Ran (2020).

Slightly different from previous research, Mikinney L's study includes international students engagement in community college and student course dropping patterns. Ghazzawi et al. (2021) exams the validity of the Community College Survey of Student Engagement (CCSSE) for international students, highlighting the inapplicability of CCSSE benchmarks in representing international student engagement. Leveraging longitudinal transcript data in Texas Community College system, McKinney et al. (2019) find that course withdrawal rates were significantly higher among students who were: male, African American, age 20–24, GED holders, academically underprepared, enrolled part-time, and had a cumulative college GPA of less than 2.0. Science, mathematics, and writing courses had high drop rates, as did Second Start and fully online courses.

Echoing with Bahr P. R.'s research about labor market for community college, V. Y. T. Liu and Fay (2020) indicate that supplementally enrolled students in community college had higher bachelor's degree attainment, and better employment outcomes than four-year college students who did not supplementally enroll. He further examines the relationship between racial/ethnic stratification and the extent of job opportunities in the labor market. Results also suggest that, while all students generally benefit from attainment of academic milestones, doing so disproportionately benefits Black and Hispanic students Lin, Fay, and Fink (2022).

Woods C. S. and Bertrand Jones T. are coauthors with Park T. and Hu S in two articles – "Reading, Writing, and English Course Pathways when Developmental Education is Optional" and "What Happens to Underprepared First-Time-in-College Students When Developmental Education is Optional?", which were discussed in above paragraphs.

Deliverable 1 Q3: Who are the top five most prolific co-authors? What are they publishing about?

```
E(g)$weight <- 1
g.c <- simplify(g)
E(g.c)$weight
```

```
##   [1] 1 1 1 1 1 1 1 2 1 1 4 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [38] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [75] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [112] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1 1 1 1 1 1 1
## [149] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [186] 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [223] 1
```

```
# this code is for Q3, the prolific co-authers
links<-as.data.frame(cbind(get.edgelist(g.c), E(g.c)$weight))

head(links,20)
```

```
##              V1               V2 V3
## 1     liu v.y.t.        fay m.p.  1
## 2       fay m.p.          lin y.  1
## 3       fay m.p.         fink j.  1
## 4        zhao k.         park t.  1
## 5        zhao k.       mokhe c.g.  1
## 6        zhao k.           hu s.  1
## 7        park t.      mokher c.g.  1
## 8        park t.       woods c.s.  2
## 9        park t.       mokhe c.g.  1
```

```
## 10       park t.             hu x.  1
## 11       park t.             hu s.  4
## 12       park t.        spencer h.  1
## 13       park t. bertrand jones t.  2
## 14       park t.       tandberg d.  1
## 15 gonzales s.m.        baier s.t.  1
## 16 gonzales s.m.      brammer e.c.  1
## 17    baier s.t.      brammer e.c.  1
## 18       naor m.        pinto g.d.  1
## 19       naor m.        davidov p.  1
## 20       naor m.       benaroch j.  1
```

```
dim(links)
```

```
## [1] 223   3
```

```
links$V3<-as.numeric(links$V3)

links<- links[order(links$V3, decreasing=T),]
```

The most prolific co-authors are Park T. and Hu S., they published 4 articles together within five years. The four papers were concentrated on the policy impact of Senate Bill 1720 on non-exempt students, course enrollment and success for underprepared first-time-in-college students. The detailed information showed above (Zhao et al. (2022); Park-Gaghan et al. (2020); Woods et al. (2019); Park et al. (2018))

The other top prolific co-authors are Woods C.S. & Park.T., Park T. & Bertrand Jones T., Woods C.S. & Hu S., Woods C.S. & Bertrand Jones T., Hu S. & Bertrand Jones T, they all published two articles together. Because this is a one-mode no direction network, we can see that Woods C.S., Park.T., Bertrand Jones T., Hu S. are all related with each other. They published two same papers together, focusing on the enrollment patterns of underprepared first-time-in-college students (Woods et al. (2019); Park et al. (2018)).

Deliverable 1 Q4: How are points 2 and 3 similar or different? In what ways?

Similarity:

(1) The articles mentioned in question 3 are all included in the question 2.
(2) Their studies are all focused on the impact of Senate Bill 1720 for Florida community colleges and course enrollment patterns for underprepared first-time-in-college (FTIC) students.
(3) In terms of code, they all need create edgelist to get the final results.

Difference:

(1) Question 2 includes more different authors and articles in this field.
(2) Question 2 contains more various research areas than question 3, varying from labor market for community college to community college students' enrollment and learning outcomes.
(3) The code for question 2 needs to construct a two-mode edgelist for author-publication, while question 3 only need a one-mode edgelist with weight.

Deliverable 2

Deliverable 2 Q1: Create a publication network wherein all relationships link (co-)authors with their respective publication. After this, describe the network, how many authors are represented, how many publications are in this network? And how many connections were established?

```
#Add publication as first column
a1 <- cbind(a$Source.title, a1)
mat <- as.matrix(a1)

edgelist_two_mode <- cbind(mat[, 1], c(mat[, -1]))

edgelist_two_mode <- edgelist_two_mode[!is.na(edgelist_two_mode[,2]), ]
g2<- graph.edgelist(edgelist_two_mode[, 2:1], directed = FALSE)  #this is from author to publication

#does this mean 146 authors, 36 journals
table(V(g2)$type)
```

```
## < table of extent 0 >
```

In the two-mode edgelist, 146 authors are represented, 36 journals are represented. Total 272 connections are established.

Deliverable 2 Q2: Who are the top five most prolific authors? What are they publishing about?

There are two authors (Park T. and Hu S.) published 4 articles in this field, one author (Bahr P. R.) published 3 articles, and 6 authors (Ngo F.; Xu D.; Mckinney L.; Woods C. S.; Fay M. P.; Bertrand Jones T.) publish 2 articles each.

Detailed publishing information illustrated in the Deliverable 1 Q2.

Deliverable 2 Q 3: What is the publication with the highest number of co-authors? What is this publication about?

```
#Transformation
links2<-as.data.frame(cbind(get.edgelist(g2)))
head(links2)
```

```
##                                       V1                                       V2
## 1            Review of Higher Education            Review of Higher Education
## 2              Community College Review              Community College Review
## 3                       Student Success                       Student Success
## 4                    Education Sciences                    Education Sciences
## 5          Research in Higher Education          Research in Higher Education
## 6 American Educational Research Journal American Educational Research Journal
```

```
#Publications
head(sort(degree(g2)[V(g2)$type==T], decreasing = T))
```

```
## named numeric(0)
```

The top six publications with the highest number of co-authors are:

(1) Community College Journal of Research and Practice - 12

CiteScore (Scopus):1.3 (2021); SJR: 0.472 (2021)

The Community College Journal of Research and Practice (CCJRP) is a peer-review journal that promotes and disseminates an increased awareness of community college issues. Topics of interest to the journal include, but are not limited to: Community college administration; Community college curriculum; Developmental education; Community college faculty; International perspectives about community colleges; Community college leadership; Community college policy; Research about community colleges; Student affairs; Community college students.

(2) AERA Open - 10

Impact Factor: 3.427; 5-Year Impact Factor: 4.166

AERA Open is a peer-reviewed, open access journal published by the American Educational Research Association (AERA). AERA Open publishes studies of education and learning in various contexts, such as early childhood, after-school, primary and secondary, and post-secondary education.

(3) Research in Higher Education - 9

Impact Factor: 2.615; 5-Year Impact Factor: 3.593

Research in Higher Education publishes studies that examine issues pertaining to postsecondary education. Among the topics of interest to the journal are: access and retention; student success; equity; faculty issues; institutional productivity and assessment; postsecondary education governance; curriculum and instruction; state and federal higher education policy; and financing of postsecondary education.

(4) Journal of Higher Education - 9

Impact Factor: 3.204 (2021); 5-year IF: 4.046 (2021)

Founded in 1930, The Journal of Higher Education publishes original research reporting on the academic study of higher education as a broad enterprise. They publish the highest quality empirical, theoretically grounded work addressing the main functions of higher education and the dynamic role of the university in society. We seek to publish scholarship from a wide variety of theoretical perspectives and disciplinary orientations.

(5) Economics of Education Review -7

CiteScore: 3.9; Impact Factor: 2.083

Economics of Education Review publishes research on education policy and finance, human capital production and acquisition, and the returns to human capital. They accept empirical, methodological and theoretical contributions, but the main focus of Economics of Education Review is on applied studies that employ micro data and clear identification strategies. Their goal is to publish innovative, cutting-edge research on the economics of education that is of interest to academics, policymakers and the public.

(6) Journal of Community Health- 7

Impact Factor: 4.371; Five-year IF: 3.425

The Journal of Community Health, a peer-reviewed publication, offers original articles on the practice, teaching, and research of community health. Coverage includes preventive medicine, new forms of health manpower, analysis of environmental factors, delivery of health care services, and the study of health maintenance and health insurance programs. Serving as a forum for the exchange of ideas and clarification, the journal features articles on projects that make a significant impact on the education of health personnel.

Deliverable 2 Q 4: How are points 2 and 3 similar or different? In what ways?

Similarity:

(1) They all yielded important information in terms of community college enrollment research.
(2) They all use two mode edgelist to get the results.

Differences:

(1) Question 2 focus on actors (authors), illustrating individual contributions to this field.
(2) Question 3 focus on the publications, demonstrating the important journals in community college field.
(3) Author selection is based on V(g2) type =F, publication selection is based on V(g2) type = T.

Deliverable 3

Deliverable 3 Q 1: Conduct the two-mode to one-mode transformation retaining co-authors.

```
a1 <- cbind(a$Source.title, a1)

mat <- as.matrix(a1)

edgelist_two_mode <- cbind(mat[, 1], c(mat[, -1]))

edgelist_two_mode <- edgelist_two_mode[!is.na(edgelist_two_mode[,2]), ]
g2<- graph.edgelist(edgelist_two_mode[, 2:1], directed = FALSE)
V(g2)$type <- V(g2)$name %in% edgelist_two_mode[,1]
```

```
mat2 <- get.incidence(g2)
```

```
## Warning in get.incidence.dense(graph, types = types, names = names, attr =
## attr): At core/misc/bipartite.c:757 : 112 edges running within partitions were
## ignored.
```

```
dim(mat2)
```

```
## [1] 146  36
```

```
mat2_to_1<-mat2%*%t(mat2)

dim(mat2_to_1)
```

```
## [1] 146 146
```

```
diag(mat2_to_1)
```

```
##           liu v.y.t.              zhao k.        gonzales s.m.
##                    1                     1                    1
##              naor m.             grosz m.              liu v.
##                    1                     1                    1
##               lin y.            vietze j.            schnee e.
##                    1                     1                    1
##               ngo f.               liu a.         griffiths r.
##                    2                     1                    1
##              sowl s.             lee m.j.             akiba d.
##                    1                     1                    1
##         fleming c.b.      dos santos l.m.        griffith c.h.
##                    1                     1                    1
##         ortagus j.c.         mathewson k.            bird k.a.
##                    1                     1                    1
##            corral d.        sorensen l.c.          ghazzawi d.
##                    1                     1                    1
##          hughes j.a.         uretsky m.c.          seidel e.j.
##                    1                     1                    1
##              park t.           schudde l.              cate r.
##                    4                     1                    1
##    chisholm-burns m.             nkana e.        matusovich h.
##                    1                     1                    1
##                xu d.          tomkin j.h.           padilla m.
##                    2                     1                    1
##          mckinney l.            bahr p.r.         cummins p.a.
##                    2                     3                    1
##           woods c.s.            thor l.m.             korah a.
##                    2                     1                    1
##         aguilar s.j.          hester s.d.        anderson d.m.
##                    1                     1                    1
##          park r.s.e.             fox h.l.         fairlie r.w.
##                    1                     1                    1
##          costello l.             krieb d.              shea p.
##                    1                     1                    1
##             fay m.p.           baier s.t.           pinto g.d.
##                    2                     1                    1
##        kurlaender m.    van herpen s.g.a.        hinojosa j.k.
##                    1                     1                    1
##           shapiro c.           mislevy j.           crain a.m.
##                    1                     1                    1
##       duckworth j.c.           de beer f.             tanner m.
##                    1                     1                    1
##          sundaram m.       castleman b.l.              hwang m.
##                    1                     1                    1
```

```
##           park h.c.          shipe s.l.          mohlman j.
##                   1                   1                   1
##         mokher c.g.  yonah meiselman a.         russ-eft d.
##                   1                   1                   1
##          chang c.f.          gillen a.            ran f.x.
##                   1                   1                   1
##      beilstein s.o.         brown j.c.            novak h.
##                   1                   1                   1
##          brown j.s.      scarafiotti c.          slate j.r.
##                   1                   1                   1
##           kim a.y.       astudillo s.           nadler m.
##                   1                   1                   1
##     goldrick-rab s.    scott-clayton j.        bidjerano t.
##                   1                   1                   1
##          mokhe c.g.        brammer e.c.          davidov p.
##                   1                   1                   1
##          stevens a.            fink j.      dias-broens a.
##                   1                   1                   1
##            gregg j.            wang s.        patrick m.e.
##                   1                   1                   1
##        edwards r.l.         mcfarlin i.     bednarczyk r.a.
##                   1                   1                   1
##            mabel z.            horn c. christensen-grant d.
##                   1                   1                   1
##    henneberger a.k.          basch c.h.               hu x.
##                   1                   1                   1
##           cooper s.          carrico c.        morphew j.w.
##                   1                   1                   1
##           abrica e.       hagedorn l.s.               hu s.
##                   1                   1                   4
##          moore g.w.          katcher j.         benaroch j.
##                   1                   1                   1
##      severiens s.e. levis-fitzgerald m.        fairlie a.m.
##                   1                   1                   1
##            smith c.            song y.       burridge a.b.
##                   1                   1                   1
##        fuentes d.g.            fera j.          spencer h.
##                   1                   1                   1
##           knight d.         herman g.l.      luna-torres m.
##                   1                   1                   1
##            mehri n.   bertrand jones t.      lunenburg f.c.
##                   1                   2                   1
##        elfring l.k.           kivel a.         meeuwisse m.
##                   1                   1                   1
##   sanders o'leary e.       abdallah d.a.           colvin g.
##                   1                   1                   1
##          carales v.         cosgrove a.            grohs j.
##                   1                   1                   1
##          dykstra e.         tandberg d.       kennison r.l.
##                   1                   1                   1
##            lee c.m.           karpf m.            ethan d.
##                   1                   1                   1
##        rezende l.f.         bolger m.s.
##                   1                   1
```

Deliverable 3 Q 2: What is the meaning of the diagonal in this one-mode transformed matrix? What is the off-diagonal telling us?

Diagonal represents the number of papers the author published.

The off-diagonal tells us the coauthor relationships, the number of time their coauthor publications.

Deliverable 3 Q 3: Compare the resulting number of authors in this dataset with the total number of authors you obtained in deliverable 1. Are these numbers the same? If they are not, elaborate on why are they different? If they are the same, also elaborate on why they are the same.

The number is different, namely, deliverable 1 has 137 authors, while deliverable 3 shows 146 authors. In my data set, it includes 9 single authors. 137 + 9 = 146. In the analysis, we should use the one-mode transformation graph (137) to see the co-author network.

Deliverable 3 Q 4: Tell us how or why these connections represented in this depiction are similar or different from those connections represented in deliverable 1? Would these network structures allow you to address different questions?

In the deliverable 3 matrix, we have four types of relationships, human to human, human to publication, publication to human, and publication to publication. Compared with matrix in deliverable 1, this matrix contains more information. In deliverable 1, we can only compute the number of publications in terms of author or journals. But in this matrix, we could see the coauthor relationships, the prolific authors, and journals.

We can look at how many different authors are covered in the same journal to determine which journals are the most popular.

Deliverable 4

Deliverable 4 Q1 Building from the first three deliverables, identify the top six most central human actors in terms of closeness, betweenness, eigenvector, and degree centralities.

```
#identify the top six most central human actors
#one-mode
#evcent(g)$vector
#plot(g)

dta1<-data.frame(ev=evcent(g)$vector, deg=degree(g),
                 bet=(betweenness(g, normalized=F)/max(betweenness(g, normalized = F))),
                 clo=closeness(g))

#two mode
#evcent(g2)$vector
#plot(g2)

dta2<-data.frame(ev=evcent(g2)$vector, deg=degree(g2),
```

```
                bet=(betweenness(g2, normalized=F)/max(betweenness(g2, normalized = F))),
                clo=closeness(g2))
```

```
#centrality table
one_mode_df <-read.table( header=T, sep='|', text='
Author|Total Degree|Eigenvector|Betweenness|Closeness
Hu S.|14|1.000|0.875|0.111
Park T.|14|1.000|0.875|0.111
Bertrand Jones T.|7|0.672|0.000|0.071
Woods C. S.|7|0.672|0.000|0.071
Tandberg D.|4|0.392|0.000|0.071
Hu X.|4|0.306|0.000|0.071' )
print(one_mode_df)
```

```
##              Author Total.Degree Eigenvector Betweenness Closeness
## 1             Hu S.           14       1.000       0.875     0.111
## 2          Park T.            14       1.000       0.875     0.111
## 3 Bertrand Jones T.            7       0.672       0.000     0.071
## 4       Woods C. S.            7       0.672       0.000     0.071
## 5       Tandberg D.            4       0.392       0.000     0.071
## 6             Hu X.            4       0.306       0.000     0.071
```

```
two_mode_df <-read.table( header=T, sep='|', text='
Author|Total Degree|Eigenvector|Betweenness|Closeness
Hu S.|4|0.608|0.673|0.008
Park T.|4|0.608|0.673|0.008
Bertrand Jones T.|2|0.395|0.055|0.006
Woods C. S.|2|0.395|0.055|0.006
Yonah Meiselman A.|1|0.250|0.000|0.005
Uretsky M. C.|1|0.250|0.000|0.005' )
print(two_mode_df)
```

```
##               Author Total.Degree Eigenvector Betweenness Closeness
## 1             Hu S.             4       0.608       0.673     0.008
## 2          Park T.              4       0.608       0.673     0.008
## 3  Bertrand Jones T.            2       0.395       0.055     0.006
## 4       Woods C. S.             2       0.395       0.055     0.006
## 5 Yonah Meiselman A.            1       0.250       0.000     0.005
## 6      Uretsky M. C.            1       0.250       0.000     0.005
```

Deliverable 4 Q2 Do you see any commonalities among these measures? If so, what does that mean?

All these four measures come from the same matrices and have some extent shared mathematical properties. They all enable the identification of movement and structure in the network through actors' position and their paths or connections.

But they also have some difference. Degree centrality assigns an importance score based simply on the number of links held by each other, which can identify actors who are likely to hold most information. Betweenness centrality measures how nodes are "bridges" between nodes in a network, illustrating the actors who influence the flow around a system. Closeness centrality calculates the shortest paths between all nodes,

suggesting the actors who are best placed to influence the entire network most quickly. Eigenvector centrality can identify nodes with influence over the whole network, not just those directly connected to it, which is a good "all-round" SNA score.

Deliverable 4 Q3 Finally, for degree centrality specify whether you are using in, out, or total and why?

In this co-authorship social network, we used total degree centrality. Because the edgelist is undirected. There are no directions between two authors.

Deliverable 5

Deliverable 5 Q1 Identify the most influential co-authors. Replicating Figure 1 (in the presentation) of the Key Actor Analysis (Optional using betweenness and eigenvector centralities).

```
cent<-data.frame(bet=betweenness(g, normalized=F)/max(betweenness(g, normalized=F)),eig=evcent(g)$vecto

rownames(cent)<-rownames(dta1)

#add residual, add coef
residuals<-lm(eig~bet,data=cent)$residuals
cent<-transform(cent,res=residuals)
coeffs<-as.data.frame(coef(lm(eig~bet,data=cent)))

summary(lm(cent$bet~cent$eig))
```

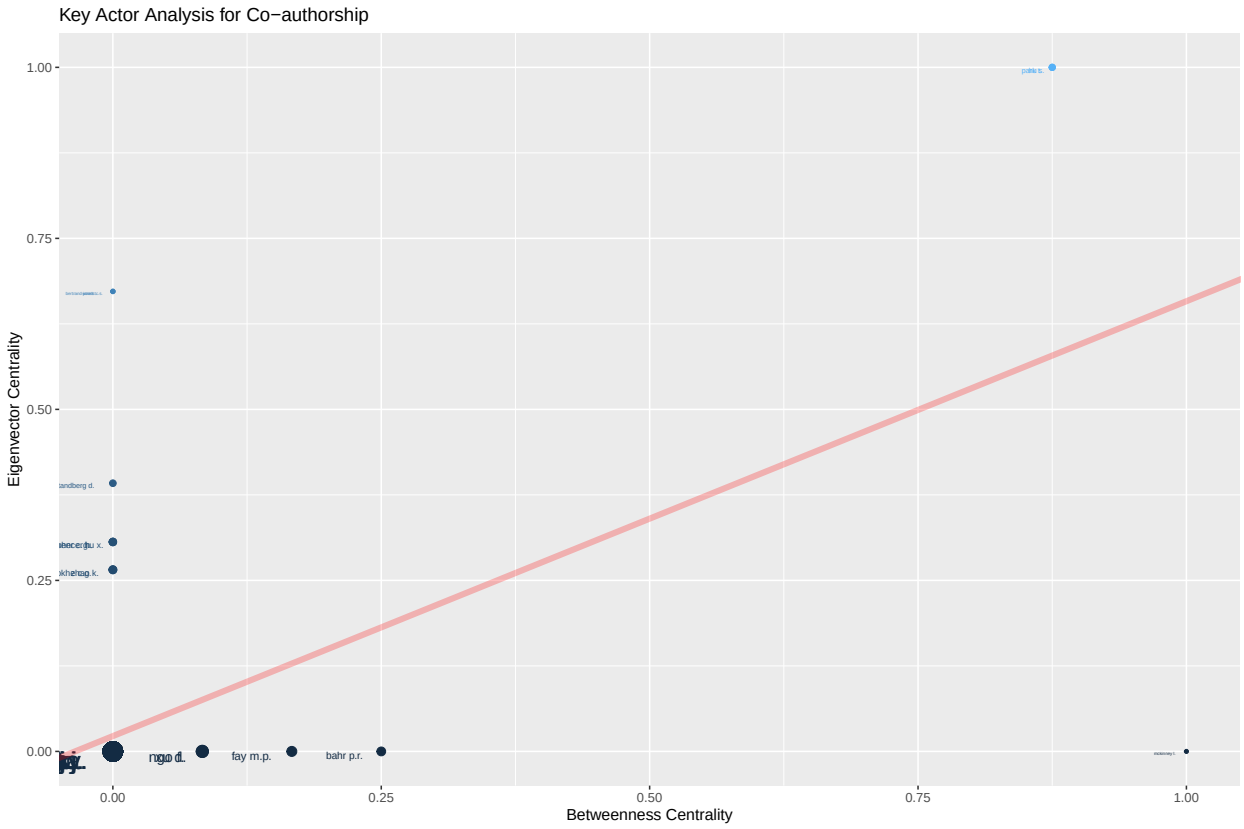
```
##
## Call:
## lm(formula = cent$bet ~ cent$eig)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.33810 -0.00561 -0.00561 -0.00561  0.99439
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.005611   0.010018   0.560    0.576
## cent$eig    0.494456   0.062853   7.867 1.04e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1139 on 135 degrees of freedom
## Multiple R-squared:  0.3143, Adjusted R-squared:  0.3093
## F-statistic: 61.89 on 1 and 135 DF,  p-value: 1.043e-12
```

```
library(ggplot2)

#code for Figure 1 eig~bet
```

```
figure1<-ggplot(cent,aes(x=bet,y=eig, label=rownames(cent),colour=eig, size=(1/abs(res))))+xlab("Betwee

figure1 +
  geom_point() +
  geom_text(hjust=1.5, vjust=1)+
  labs(title="Key Actor Analysis for Co-authorship") +
  geom_abline(intercept = coeffs[1,], slope = coeffs[2,],colour = "red", size = 2,alpha=.25) +
  theme(legend.position = "none")
```



Deliverable 5 Q2 Replicate Figure 2 (the weighted sociogram in the presentation), making sure that only the top 5% of influential actor names are shown.

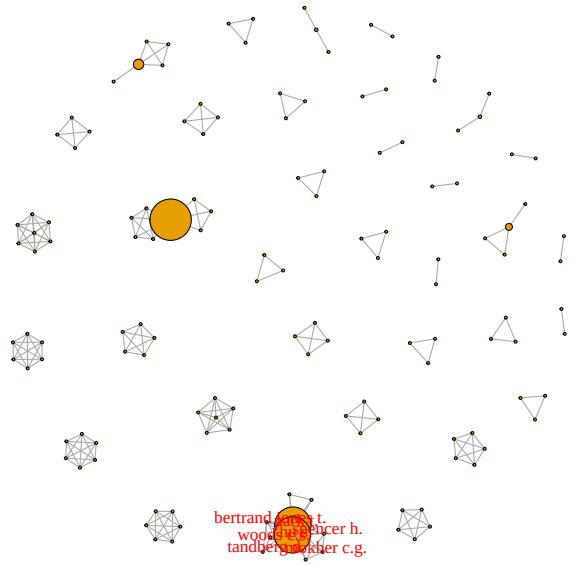
```
#code for figure 2
set.seed(47)
l<-layout.fruchterman.reingold(g, niter=5000)
V(g)$name<-rownames(dta1)
V(g)$size<-abs((cent$bet)/max(cent$bet))*15#The divisor is the highest betweenness
nodes<-V(g)$name

#set 5% threshold
x<-quantile(cent$eig, .95)
nodes[which(abs(cent$eig)<(x))]<-NA

table(is.na(nodes))
```

```
##
## FALSE  TRUE
##     7   130
```

```
plot(g,layout=l,vertex.label=nodes, vertex.label.dist=0.25,
     vertex.label.color="red",edge.width=1)
title(main="Key Actor Analysis of Co-authorship", sub="Key actors weighted by BC, names are top 5% EC",
```



Deliverable 5 Q3 Finally Provide an analysis of Figures 1 and 2 building upon the findings of each figure as shown in class. Specifically, address whether the use of both figures helped you gain a better understanding.

As is shown in figure 1, Hu S. and Park T. have the highest betweenness centrality and eigenvector centrality. The plot shows that they are doing better in eigenvector centrality than what they were expected to perform based on their performance in betweenness centrality. Bahr P. R. and Fay M. P. have higher betweenness centrality than others. Besides, Bertrand Jones T. and Woods C. S. have higher eigenvector centrality than others.

I also conducted a summary of the linear regression between eigenvector centrality and betweenness centrality. As is shown in the above summary table, the coefficient is 0.494, with a significance level of p-value less than 0.001, indicating that betweenness centrality has significant impact on eigenvector centrality. And the R square of this model is 0.314, suggesting that betweenness centrality could explain 31.4% of the eigenvector centrality.

As is shown in figure 2, the top 5% influential actors include Park T., Hu S., Woods C. S.; Spencer H.; Mokher C. G.; Bertrand Jones T.; and Tandberg D.

Combining the last three assignments, we counted the top five most prolific authors (Park T.; Hu S.; Bahr P. R.; Ngo F.; Xu D.) and top five most prolific co-authors (Park T.; Hu S.; Woods C.S.; Bertrand Jones T.; Fay M. P.), which are slightly different from top 5% influential actors. The second graph helps us to filter the most influential authors (in terms of betweenness centrality), and the intuitive visualization helps us to locate the key authors and reduce the influence of noisy information.

Reference

Bahr, Peter Riley. 2019. "The Labor Market Returns to a Community College Education for Noncompleting Students." *The Journal of Higher Education* 90 (2): 210–43.

Cummins, Phyllis A, J Scott Brown, Peter Riley Bahr, and Nader Mehri. 2019. "Heterogeneity of Older Learners in Higher Education." *Adult Learning* 30 (1): 23–33.

Fairlie, Robert W, and Peter Riley Bahr. 2018. "The Effects of Computers and Acquired Skills on Earnings, Employment and College Enrollment: Evidence from a Field Experiment and California UI Earnings Records." *Economics of Education Review* 63: 51–63.

Ghazzawi, Dina, Lyle McKinney, Catherine Horn, Andrea Backsheider Burridge, and Vincent Carales. 2021. "Reconceptualizing Student Engagement: Investigating the Validity of CCSSE Benchmarks as Indicators of International Student Engagement." *Journal of International Students* 11 (2): 436–58.

Lin, Yuxin, Maggie P Fay, and John Fink. 2022. "Stratified Trajectories: Charting Equity Gaps in Program Pathways Among Community College Students." *Research in Higher Education*, 1–27.

Liu, Vivian Yuen Ting, and Margaret P Fay. 2020. "Does Taking a Few Courses at a Community College Improve the Baccalaureate, STEM, and Labor Market Outcomes of Four-Year College Students?"

Liu, Vivian, and Di Xu. 2022. "Happy Together? The Peer Effects of Dual Enrollment Students on Community College Student Outcomes." *American Educational Research Journal* 59 (1): 3–37.

McKinney, Lyle, Heather Novak, Linda Serra Hagedorn, and Maria Luna-Torres. 2019. "Giving up on a Course: An Analysis of Course Dropping Behaviors Among Community College Students." *Research in Higher Education* 60 (2): 184–202.

Ngo, Federick, and Samantha Astudillo. 2019. "California DREAM: The Impact of Financial Aid for Undocumented Community College Students." *Educational Researcher* 48 (1): 5–18.

Ngo, Federick, and Juanita K Hinojosa. 2022. "Broadened Possibilities: Undocumented Community College Student Course Enrollment After the California DREAM Act." *AERA Open* 8: 23328584221091276.

Park, Toby, Chenoa S Woods, Shouping Hu, Tamara Bertrand Jones, and David Tandberg. 2018. "What Happens to Underprepared First-Time-in-College Students When Developmental Education Is Optional? The Case of Developmental Math and Intermediate Algebra in the First Semester." *The Journal of Higher Education* 89 (3): 318–40.

Park-Gaghan, Toby J, Christine G Mokher, Xinye Hu, Hayley Spencer, and Shouping Hu. 2020. "What Happened Following Comprehensive Developmental Education Reform in the Sunshine State? The Impact of Florida's Developmental Education Reform on Introductory College-Level Course Completion." *Educational Researcher* 49 (9): 656–66.

Woods, Chenoa S, Toby Park, Shouping Hu, and Tamara Bertrand Jones. 2019. "Reading, Writing, and English Course Pathways When Developmental Education Is Optional: Course Enrollment and Success for Underprepared First-Time-in-College Students." *Community College Journal of Research and Practice* 43 (1): 5–25.

Xu, Di, and Florence Xiaotao Ran. 2020. "Noncredit Education in Community College: Students, Course Enrollments, and Academic Outcomes." *Community College Review* 48 (1): 77–101.

Zhao, Kai, Toby J Park-Gaghan, Christine G Mokhe, and Shouping Hu. 2022. "Examining the Impacts of Florida's Developmental Education Reform for Non-Exempt Students: The Case of First-Year English and Math Course Enrollment and Success." *Community College Review* 50 (2): 171–92.