discourseGT: An R package to analyze discourse networks in educational contexts

Joshua P. Le[1], **Albert Chai**[1], **Andrew S. Lee**[2], **Jitarth Sheth**[1], **Kevin Banh**[1], **Priya Pahal**[1], **Katherine Ly**[1], **Qi Cui**[1], **and Stanley M. Lo**[1]

1 University of California, San Diego **2** University of California, Los Angeles

Summary

Student discussions in the classroom are important components of the learning process. Research methods for analyzing discourse in these educational contexts are predominantly qualitative. `discourseGT` is an R package that adapts graph theory to analyze discourse networks of students collaborating in small groups. This software package takes data on the sequential student talk-turns in a classroom discussion and produces statistics and generates plots based on graph theory and additional parameters. Overall, `discourseGT` provides new features that can yield insight on the dynamics of student discussions relevant to education researchers.

Statement of Need

Many contemporary applications and software packages are optimized for large-scale networks. For example, `igraph` (Csardi & Nepusz, 2006), `network` (Butts, 2021), and `sna` (Butts, 2020) were developed to analyze social media networks (Jones, Bond, Bakshy, Eckles, & Fowler, 2017), epidemiological networks (Christakis & Fowler, 2011), and political networks (Hobbs, Burke, Christakis, & Fowler, 2016), respectively. In contrast, discourse networks in educational contexts are substantially smaller, typically with only 3-8 students (Wagner & González-Howard, 2018). Consequently, certain parameters that are relevant for these larger networks are not necessarily applicable, and analysis of discourse networks demands additional parameters beyond what is available in graph theory (Chai, Le, Lee, & Lo, 2019; Lou, Abrami, & d'Apollonia, 2001).

Usage

`discourseGT` comes equipped with example data. We use this dataset to demonstrate the utility of `discourseGT` in examining discourse networks.

Preparing an igraph Object

An `igraph` object is the core input to many of the modular analytical functions offered in `discourseGT`. Prior to generating an `igraph` object, a weighted edge list needs to be generated from imported raw data, structured as two columns containing sequential nodes or individual students who start or continue a discussion episode (Chai, Le, Lee, & Lo,

2019). This is addressed by the `tabulate_edges` function. By default, the weight of an edge is defined as the number of times an edge has occurred between two nodes. Weights can be redefined based on other available criteria, but this must be done manually.

```
# Calculate the weighted edge list
tabEdge <- tabulate_edges(data, iscsvfile = FALSE, silentNodes = 0)
# Check the weighted edge list
head(tabEdge$master)
```

```
##   source target weight
## 1      1      1      8
## 2      2      1     25
## 3      3      1     49
## 4      4      1     75
## 5      1      2     28
## 6      3      2     11
```

Generation of an `igraph` object is handled by the `prepareGraphs` function, which requires the following information:

- The variable that stores the weighted edge list
- The title of the project. Default: `null`
- Is the graph weighted? Default: `TRUE`

```
prepNet <- prepareGraphs(
  tabEdge,
  project_title = "Sample Data 1",
  weightedGraph = TRUE
)
```

The graph settings specified by `prepareGraphs` will influence the analytical output of downstream functions.

Running Graph Theory Analysis

discourseGT offers graph theory-based analytics via two separate functions: `coreNetAnalysis()` and `subgroupsNetAnalysis()`.

`coreNetAnalysis()` will perform core graph theory operations, such as the counting number of nodes and edges and calculating edge weights, average graph degree, centrality, and other graph theory parameters (Chai et al., 2019).

```
coreNet <- coreNetAnalysis(prepNet)
```

`subgroupsNetAnalysis()` utilizes the Girvan-Newman algorithm to detect subgroups within the overall network (Girvan & Newman, 2002), such that:

```
subNet <- subgroupsNetAnalysis(
  prepNet, raw_input = data,
  normalized = TRUE
)
```

Generating Summaries

While it is possible to display the generated `igraph` object, core network statistics, and subgroup statistics as separate outputs, it can be helpful to view them as an overall summary of a network's graph theory analytics. Furthermore, combining all of these outputs into a single variable is a necessary step in exporting them as a single text file. The `summaryNet()` function will combine the outputs from `prepareGraphs()`, `coreNetAnalysis()`, and `subgroupsNetAnalysis()` as such:

```
summaryData <- summaryNet(
  netintconfigData = prepNet,
  coreNetAnalysisData = coreNet,
  subgroupsNetAnalysisData = subNet,
  display = TRUE
)
```

Basic Visualization

`discourseGT` offers several methods to visualize networks. For a basic network graph, `basicPlot()` is used, which offer parameters that modify the plotting algorithm, edge curvature, arrow size, and edge weight scaling.

The default plotting algorithm of `basicPlot()` is Fruchterman-Reingold, denoted by `0` (Fruchterman & Reingold, 1991). This is typically the best option to use because it attempts to minimize edge intersections in the final plot, improving readabiliy. Other projections include Kamada-Kawai (Kamada & Kawai, 1989) and Reingold-Tilford (Reingold & Tilford, 1981), denoted by `1` and `2`, respectively.

```
basicPlot(prepNet, graph_selection_input = 0, curvedEdgeLines = TRUE,
          arrowSizeMultiplier = 2, scaledEdgeLines = TRUE,
          scaledMin = 1, scaledMax = 10)
```

Sample Data 1
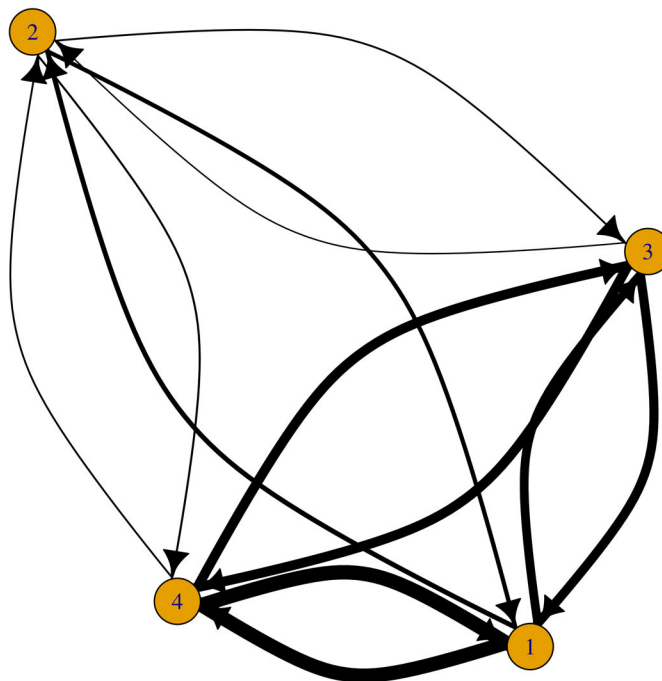
Running Non-Graph Theory Analysis

discourseGT does not require an `igraph` object to produce a non-graph theory (NGT) analysis. Rather, the `plotNGTData()` function utilizes the two-column raw data to generate its output. Additionally, it requires the duration of the conversation (in minutes) and the number of silent nodes (i.e. students who did not speak at all) in the discourse network.

```
plotNGTData(data = data, convoMinutes = 90,
            iscsvfile = FALSE, silentNodes = 0)
```

Conclusions

We developed an R package `discourseGT` that considers student discourse as a network and quantitatively examines the dynamics of small-group discussions. This paper offers a step-by-step case example that contextualizes the workflow with educational data. Details about usage and more elaborate examples are hosted online at the Comprehensive R Archive Network (https://cran.rstudio.com/web/packages/discourseGT/index.html).

Acknowledgements

References

Butts, C. T. (2020). *Sna: Tools for social network analysis.* Retrieved from https://CRAN.R-project.org/package=sna

Butts, C. T. (2021). *Network: Classes for relational data.* The Statnet Project (http://statnet.org). Retrieved from http://CRAN.R-project.org/package=network

Chai, A., Le, J. P., Lee, A. S., & Lo, S. M. (2019). Applying graph theory to examine the dynamics of student discussions in small-group learning. *CBE - Life Sciences Education*, *18*. doi:10.1187/cbe.18-11-0222

Christakis, N. A., & Fowler, J. H. (2011). *Connected: The surprising power of our social networks and how they shape our lives – how your friends' friends' friends affect everything you feel, think, and do.* Little, Brown; Company.

Csardi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, 1695. Retrieved from http://igraph.org

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software - Practice and Experience*, *21*, 1129–1164. doi:10.1002/spe.4380211102

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, *99*, 7821–7826. doi:10.1073/pnas.122653799

Hobbs, W. R., Burke, M., Christakis, N. A., & Fowler, J. H. (2016). Online social integration is associated with reduced mortality risk. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 12980–12984. doi:10.1073/pnas.1605554113

Jones, J. J., Bond, R. M., Bakshy, E., Eckles, D., & Fowler, J. H. (2017). Social influence and political mobilization: Futher evidence from a randomized experiment in the 2012 u.s. Presidential election. *PLoS ONE*, *12*. doi:10.1371/journal.pone.0173851

Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, *31*(1), 7–15. doi:https://doi.org/10.1016/0020-0190(89)90102-6

Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research*, *71*(3), 449–521. doi:10.3102/00346543071003449

Reingold, E. M., & Tilford, J. S. (1981). Tidier drawings of trees. *IEEE Transactions on Software Engineering*, *7*(2).

Wagner, C. J., & González-Howard, M. (2018). Studying discourse as social interaction: The potential of social network analysis for discourse studies. *Educational Researcher*, *47*(6), 375–383. doi:10.3102/0013189X18777741