Graph Clustering for the High School Classroom

EMILIE HOGAN*, GABRIELA RADU†

*Pacific Northwest National Laboratory,
†Liberty Ranch High School, STEM Teacher and Researcher (STAR) program

Joint Mathematics Meetings, January 17, 2014
MAA Session on Instructional Approaches to Increase Awareness of the Societal Value of Mathematics

Outline

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

STAR

- STAR program
- Motivating example for graph clustering
- Lesson plan
- Observations and student work

STAR program

- Mission
  - STAR aims to produce excellent K-12 STEM teachers by providing aspiring teachers with opportunities to do authentic STEM research and helping them translate their research experience into classroom practice
- About the program
  - STAR has three major components:
    - summer research internships,
    - education workshops during the internship,
    - opening and closing conferences
  - Expectations
    - Research presentation
    - Develop lesson plan based on research

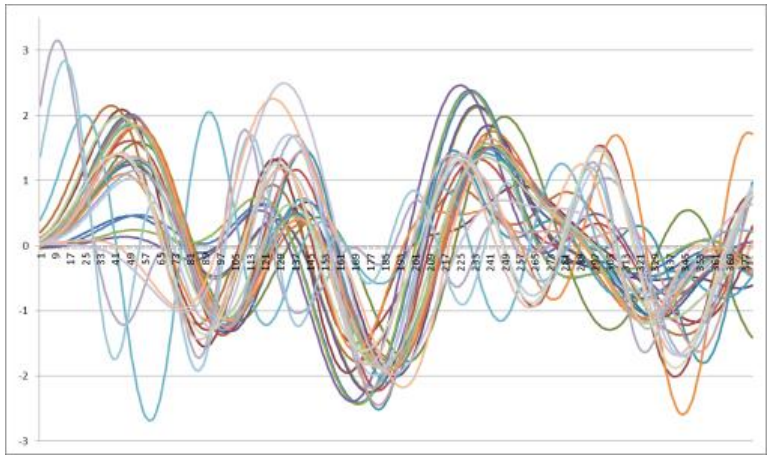
Outline

- STAR program
- Motivating example for graph clustering
- Lesson plan
- Observations and student work

Graph clustering in the real world

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

- Power grid scenario:
  - Power companies are responsible for generators within their own system
  - There are connections across company lines, systems are not isolated from each other
  - To have an accurate model of the system the companies will:
    - fully model their own generators
    - try to find a reduced model of the external system
- Solution:
  - Group generators together based on their phasor measurement unit data
  - Behavior change after system disturbance, new reduced model may be needed

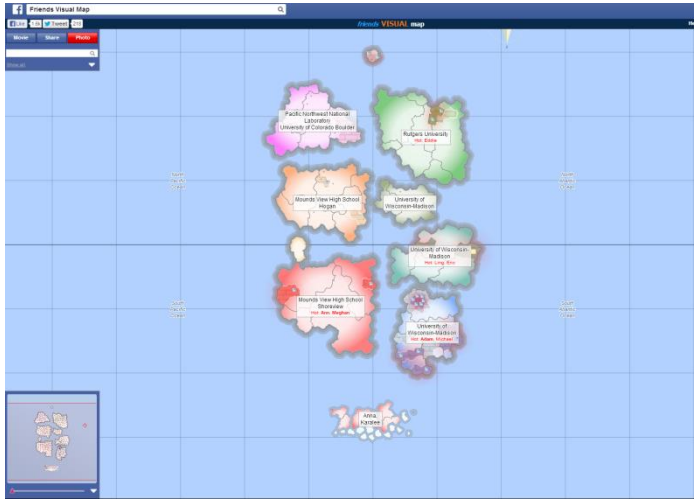


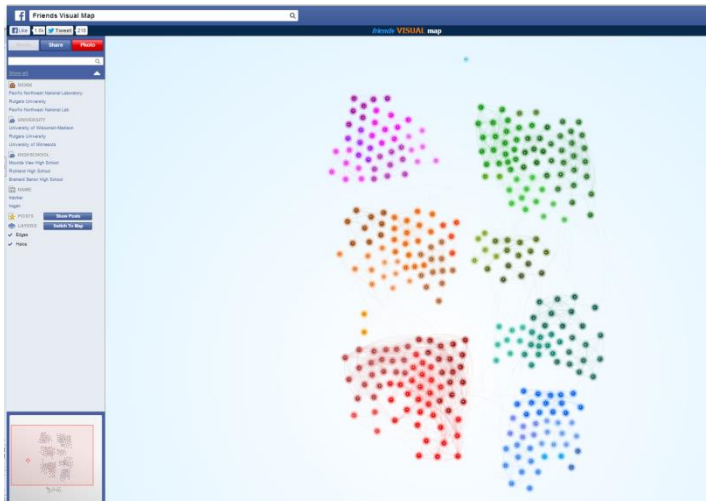
Making the graph

- Given time series data, consider each curve (representing data from one generator) as a vector of real numbers (one value per time point)
- For each pair of generators we find the *distance* between their time series vectors
  - Distance can be Euclidean or a pseudo-distance based on the discrete Fourier transform (DFT) of the vector
- Given all pairwise distances, we create a nearest-neighbor graph where vertices are the generators
  - k-NN : for each generator, connect it via an edge to the k closest generators, for some given k
  - ϵ-NN : for each generator, x, connect it to generator y if $dist(x, y) < \epsilon$
- From the graph we use graph clustering techniques:
  - Spectral, evolutionary algorithm, SVD on the adjacency matrix, etc.

Moving toward a lesson plan

- Students may not be motivated by power grid problems, but graph clustering shows up elsewhere in the real world
- Gabriela chose to use Facebook as a motivating example for the students
  - Relevant to high school students
  - Students know about it
  - Facebook (or apps within Facebook) actually use clustering!





Outline

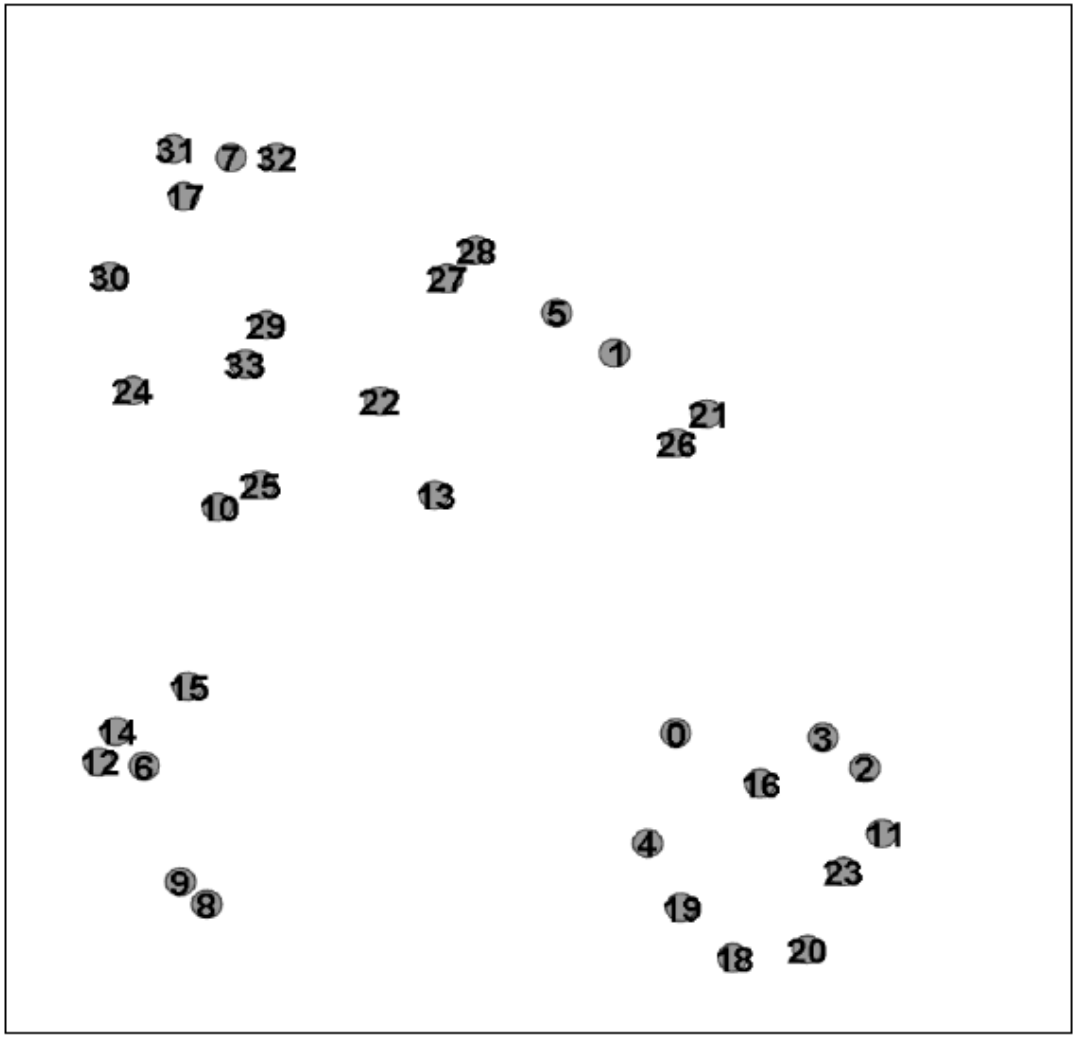
- STAR program
- Motivating example for graph clustering
- Lesson plan
- Observations and student work

Lesson Plan – Hook

- Ever wonder how Facebook “suggests” friends?
  - Why do they suggest that you be 'friends' with someone when you only have 5 friends in common?
  - If you have 556 friends, does 5 friends in common mean anything?
  - If you have 74 friends, does 5 friends in common mean anything?
- Graph clustering is one way a computer system could predict with whom you would be friends
  - If you have hundreds of people (vertices) and some are closer (distance) than others, does that mean they belong in the same group?
  - And if they do belong in the same group, is there one person that we can use as the main 'friend'?

Lesson Plan – materials

- Students were given this worksheet along with a compass
- Students were told that the vertices represent people, and that some have common traits
- Based on their “distance” from each other, form clusters and identify what they could represent
- They were asked to form clusters with these points



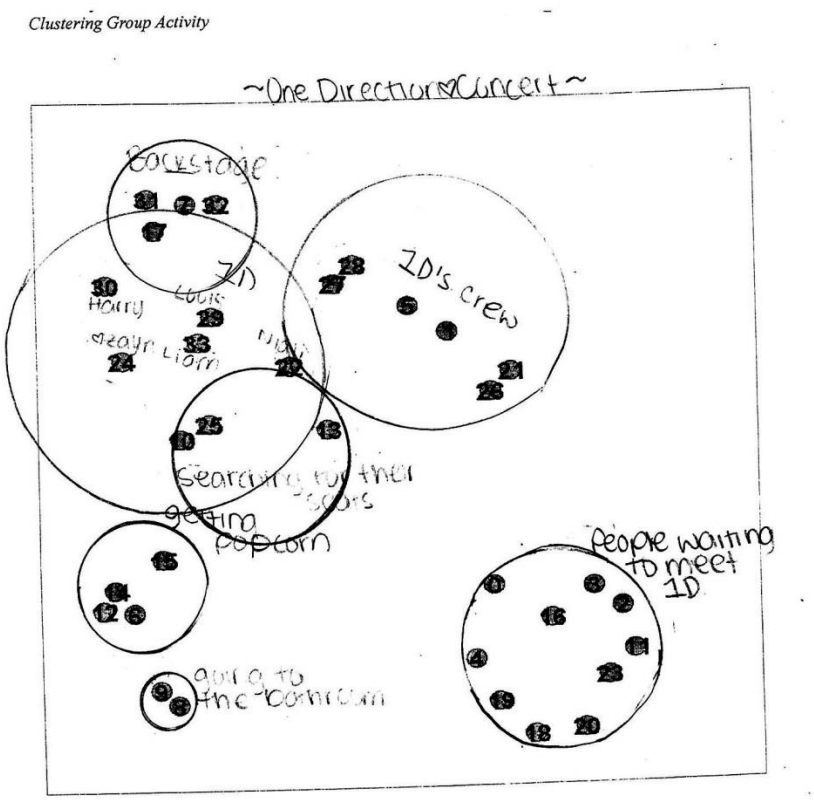
Lesson procedure

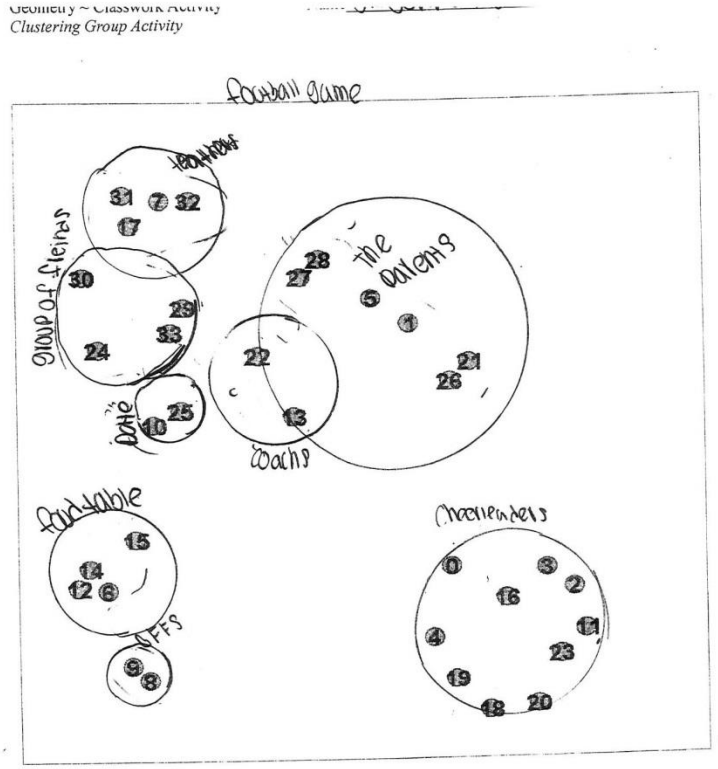
- Groups of students determined clusters and came up with traits for each group
- Class discussed as a group:
  - Number of clusters
  - Pros/Cons of many or few clusters
  - Can you start with few clusters and make more? How?
  - What if you used a compass to form circular clusters? How should you draw the circles? With a vertex in the center or no?

Outline

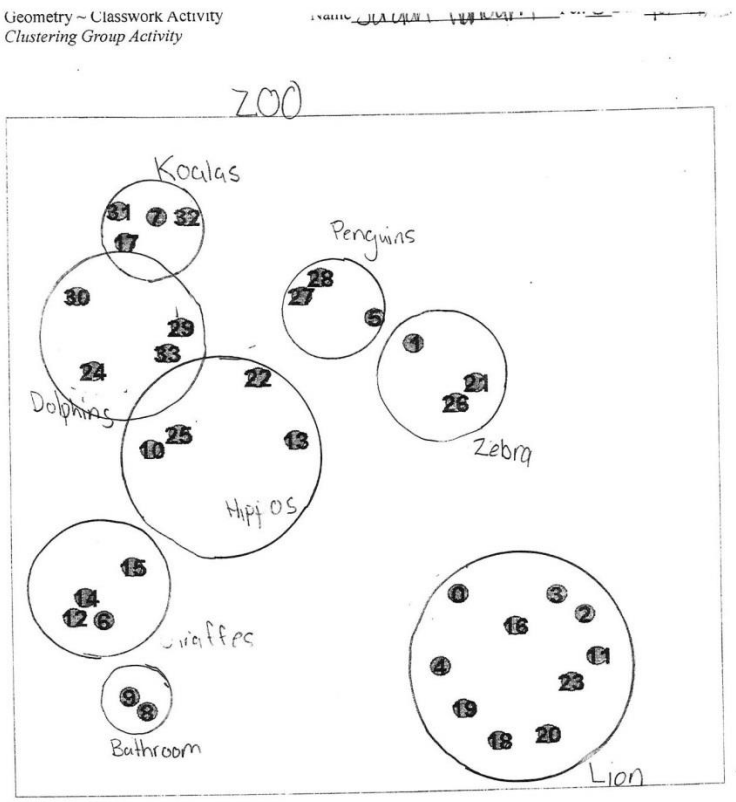
- STAR program
- Motivating example for graph clustering
- Lesson plan
- Observations and student work

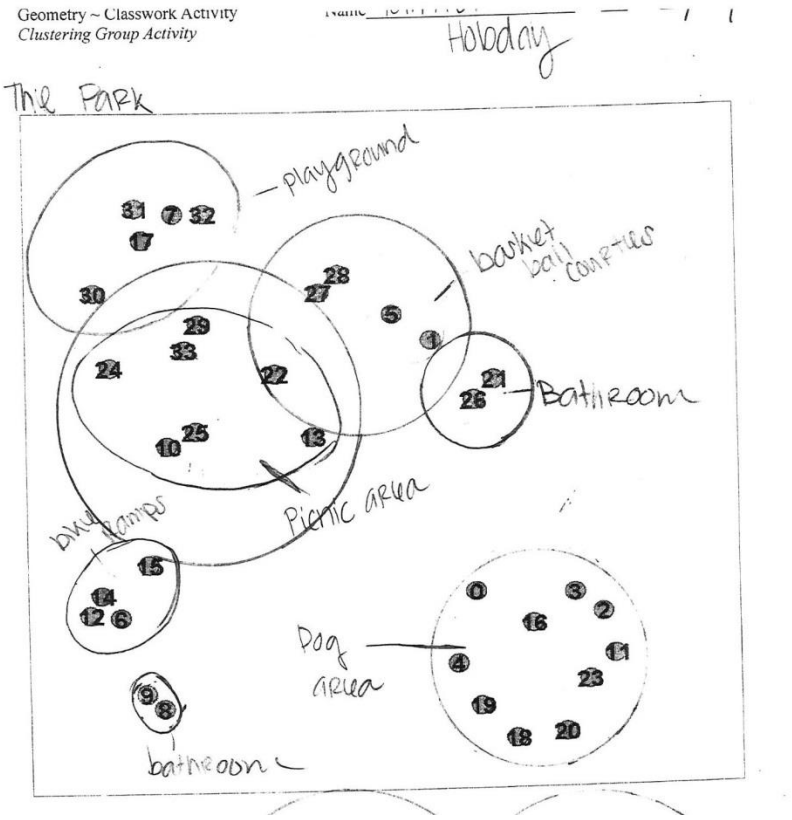
Example work



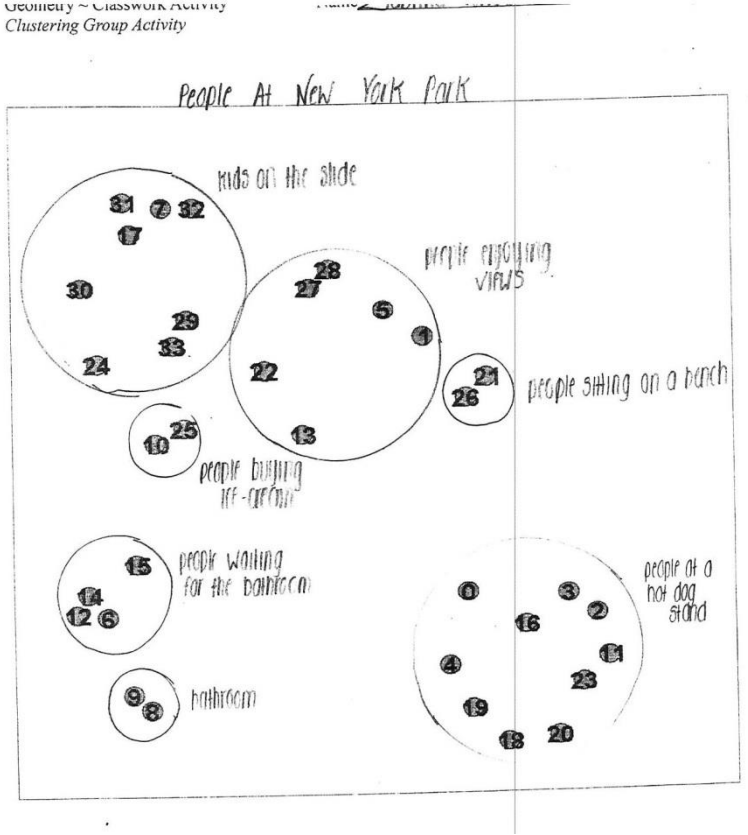


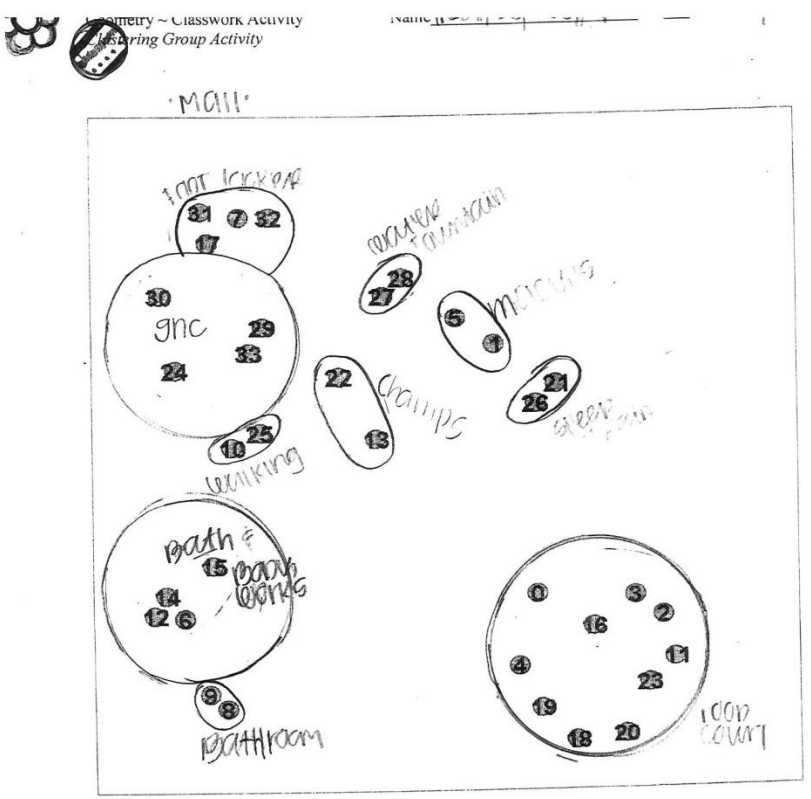
Example work (cont.)





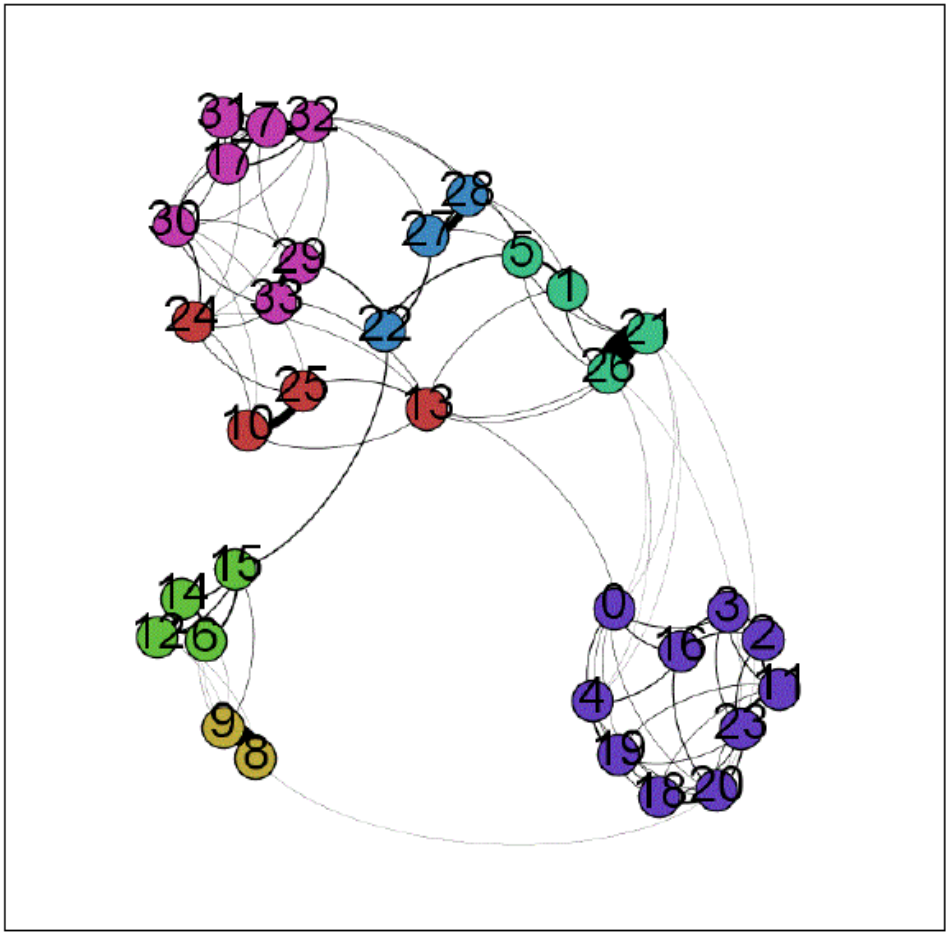
Example work (cont.)





“Real” results

- A clustering based on the real world application of clustering generators in the power grid
- Everybody got the purple, green, and the yellow clusters
- Less consensus on the upper portion (blue, red, pink, aqua clusters)
  - As many as 7 clusters and as few as 4



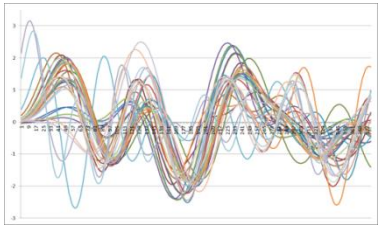
Teacher observations

- Students were initially confused, but the Facebook explanation helped
- What is a benefit for having a smaller (larger) amount of clusters? The down side?
  - Down side of more clusters – they had to write more reasons as to why things were grouped together
  - Writing more causes more work, they stayed with 5-7 clusters
- Could we tell from the number of vertices how many clusters we will get?
  - Solid NO
- What would be the smallest/largest number of clusters? What would they look like?
  - They were able to say that the largest cluster would produce only one characteristic and the smallest clusters would be each dot.

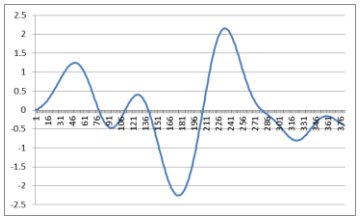
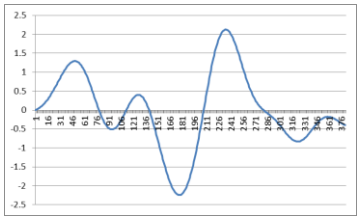
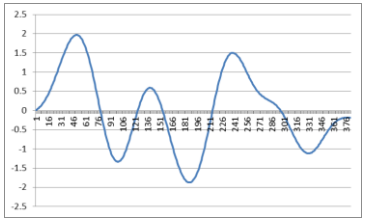
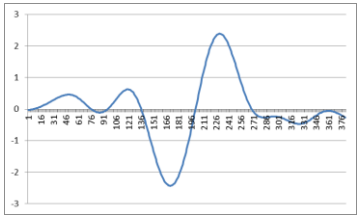
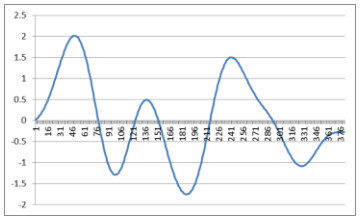
Teacher observations (cont.)

- Can you have overlapping circles? How big should the circles be? Where is the center of the circle? What if one vertex is in two circles?
  - Drawing with the compass was their biggest issue.
  - Could not figure out how to draw circles that did NOT have overlapping vertices.
  - Had to choose and/or modify their qualities to accommodate for the overlapping vertices. (Teacher was very impressed by this)
  - Majority choose a non-vertex as the center of their circles because they knew that their circle had to accommodate all their dots.
  - The center of the vertices was almost a “duh” thing for them.
  - They thought my question of where should we put the center of the circle was kind of silly.
- Student take-aways
  - Math problems don’t always have a “right answer”
  - Justifying their answers was difficult, not a common practice

Lessons learned



- Next time, have them actually look at time series because I think they could handle converting the curves into vertices.
- Or, have each individual generators' time series on a graph and have them sort the graphs based on the similarities and see if that matches with the Gephi clusters.
  - Gephi is a graph visualization tool which has a built in graph clustering algorithm
  - This would be a good lesson for a pre-Calculus class since they have an idea of what sine and cosine graphs look like.

Summary

- High school math teacher Gabriela Radu worked with me on the real graph clustering application – power grid
- She took that experience, as a STEM Teacher and Researcher (STAR) Program intern, and turned it into a classroom lesson relevant to the students
- Students learned one way mathematics research can be useful
- The lesson was successful and the teacher learned valuable lessons that she can use in the future

Acknowledgements

- The authors acknowledge the contributions of Mahantesh Halappanavar and Eduardo Cotilla-Sanchez to this work.
- This work was supported in part by the Applied Mathematics Program of the Office of Advance Scientific Computing Research within the Office of Science of the U.S. Department of Energy (DOE).
- Pacific Northwest National Laboratory (PNNL) is operated by Battelle for the DOE under Contract DE-AC05-76RL01830.