

# Community Detection in Video Games Sales Dataset

Chanel Thorpe

## Overview:

My project performs community detection analysis on a dataset containing information about video game sales. It utilizes the Label Propagation Algorithm to identify the communities. The Label Propagation Algorithm is a semi-supervised learning algorithm used primarily for community detection in graphs. I decided I wanted to look at two specific columns (or groups of columns) and chose to identify communities based on genre and sales data. For the genre community, I am essentially showing the association between game genres and the platforms they're available on. For the sales community, I am showing the relationship between game genres and their corresponding sales across different platforms. Further explanation with the output will be mentioned later in this report.

## Dependencies:

'csv': I used this external crate for reading and parsing the csv file.

'serde' and 'serde\_derive': I used this external crate for serializing and deserializing data, which enabled the integration with csv parsing.

'petgraph': I used this external crate to provide graph data structure and algorithms

## VideoGame Struct:

I defined the VideoGame struct only mentioning the columns that I would be using throughout my code. The columns I didn't plan to work with and therefore left out were rank, name, year, publisher, other\_sales, and global\_sales. I only chose to work with genre, platform, North America sales, European sales, and Japan sales. I was able to use these only after cleaning the first vgsales.csv file (which is included in this file) and using the cleaned version.

## Community Detection Module:

'detect\_communities': This function reads the dataset from a CSV file specified by the 'file\_path' parameter. It constructs adjacency lists for genre-based and sales-based graphs, where each genre or platform is a node, and edges represent relationships between them. It then applies

the Label Propagation Algorithm to detect communities based on these graphs and prints the results.

‘label\_propagation’: This function implements the Label Propagation Algorithm to detect communities in a genre-based graph. It starts from each node, propagating labels to neighboring nodes until convergence, forming communities.

‘label\_propagation\_sales’: This function performs label propagation on a sales-based graph, considering the sales data as weights on the edges. It aims to identify communities based on sales patterns among different genres and platforms.

### Testing:

I decided to test the two label propagation functions to see how accurate they were:

‘test\_label\_propagation’: Tests the label propagation algorithm for genre-based community detection.

‘test\_label\_propagation\_sales’: Tests the algorithm for sales-based community detection.

```
● Finished test [unoptimized + debuginfo] target(s) in 0.09s
  Running unittests src/main.rs (target/debug/deps/project_
file-13362c6ab2e3332b)

running 2 tests
test tests::test_label_propagation_sales ... ok
test tests::test_label_propagation ... ok

test result: ok. 2 passed; 0 failed; 0 ignored; 0 measured; 0
filtered out; finished in 0.00s

* Terminal will be reused by tasks, press any key to close i
t.
□
```

## Output and Explanation:

```
● Compiling project-file v0.1.0 (/Users/chanelthorpe/GitHub/D
S210-Final-Project/project-file)
Finished dev [unoptimized + debuginfo] target(s) in 1.22s
Running `target/debug/project-file`
Communities based on Genre:
Community 0: {"PS4", "Wii", "PS3", "PS", "X360", "XOne", "GEN",
"Fighting", "GC", "NES", "PSP", "GBA", "XB", "PS2", "N64", "
WiiU", "SAT", "SNES", "3DS"}
Community 1: {"DC", "2600", "GB", "PSV", "DS", "SCD", "Platform"}
Community 2: {"Adventure", "PC"}
Community 3: {"Action"}
Community 4: {"Strategy"}
Community 5: {"Shooter"}
Community 6: {"Puzzle"}
Community 7: {"Racing"}
Community 8: {"Simulation"}
Community 9: {"Role-Playing"}
Community 10: {"Sports"}
Community 11: {"Misc"}
Communities based on Sales:
Community 0: {"NES", "GC", "PC", "PS", "DC", "GBA", "Wii", "XOne", "DS", "PS4", "GB", "Adventure", "PS3", "PS2", "X360"}
Community 1: {"PSP", "SNES", "N64", "Role-Playing", "XB", "PSV", "3DS"}
Community 2: {"Fighting", "SAT", "WiiU", "GEN"}
Community 3: {"2600", "Shooter"}
Community 4: {"Strategy"}
Community 5: {"Misc"}
Community 6: {"Action"}
Community 7: {"Simulation"}
Community 8: {"Sports"}
Community 9: {"Puzzle"}
Community 10: {"Platform", "SCD"}
Community 11: {"Racing"}
* Terminal will be reused by tasks, press any key to close it.
□
```

### Communities based on Genre:

Each community represents a group of video game platforms that share similar genres.

Community 0: Contains multiple platforms including PlayStation, Xbox, Nintendo consoles, but only one genre, "Fighting". This basically explains that across many platforms, the genre that is most closely associated with these platforms is fighting. This isn't surprising at all but it is quite interesting to see how popular the genre is regardless of platform for the most part.

Community 1: Includes a slightly more diverse and older representation of consoles/platforms like Game Boy (GB) and Atari 2600 (2600). The genre in this community is "Platform". The "Platform" genre is basically a description of the gameplay where players control characters and their movements across platforms in the game. This is quite interesting and it could be confirmed that this community is basically grouping "classics" together which would be both the platform (console type) and genre.

Community 2: The only labels in this community are "Adventure" and "PC". Most of the platforms within this community have the adventure or pc label.

Communities 3-11: Each represents a single genre, such as Action, Strategy, Shooter, Puzzle, Racing, Simulation, Sports, Role-Playing, and Miscellaneous (Misc). These communities group platforms that predominantly host games of these respective genres and do not have any specific associations with other genres.

### **Communities based on Sales:**

Each community represents a group of sales data across different platforms.

Community 0: Interestingly enough, this community holds various platforms with high sales across various genres, including NES, PC, PlayStation (PS, PS2, PS3, and PS4), Xbox, and others. Something else to notice is that the genre "Adventure" is also included here. This suggests that across all of these platforms, the genre closely related with all of them is "Adventure". This community, I think, is basically showing that these were the best sellers whenever they came out.

Community 1: This community shows the platforms with moderate sales. This includes platforms that were mostly hand-held consoles, hosting games primarily of the Role-Playing genre. Consoles like PSP, SNES, N64, DS, PSV, and 3DS are included in this community.

Community 2: This community primarily focuses on fighting and consoles such as SAT, WiiU, and GEN.

Communities 3-11: Each represents platforms with sales primarily focused on specific genres, such as Shooter (Community 3), Strategy (Community 4), Miscellaneous (Misc) (Community 5), and so on.

### **Conclusion and Future Implementation:**

Overall I think my code does a good job in looking at the relationship between genre and console and then looking at the effect of genre and console on sales. I think I could've done a bit more in looking specifically at sales in each nation but I am still happy with the results. I think for future implementations I could look at the year it was released, for example, to draw more conclusive inferences. When I was talking about community 1 under Communities based on genres, I inferred that that community is basically a grouping of all the classics both in the platform sense and the genre sense. But I don't have a way to actually check this guess because I didn't look at the years of releases. Another possible thing I could take a look at is the publisher to see who is most successful in the game world (possibly incorporating a year here too).