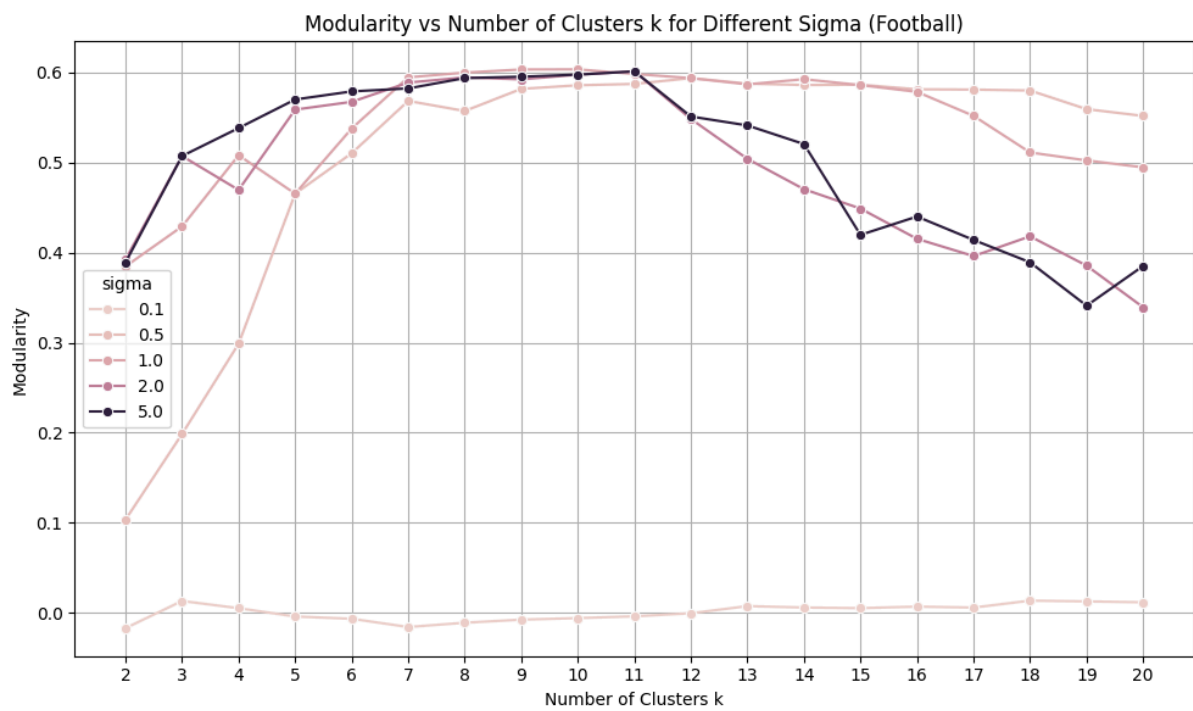
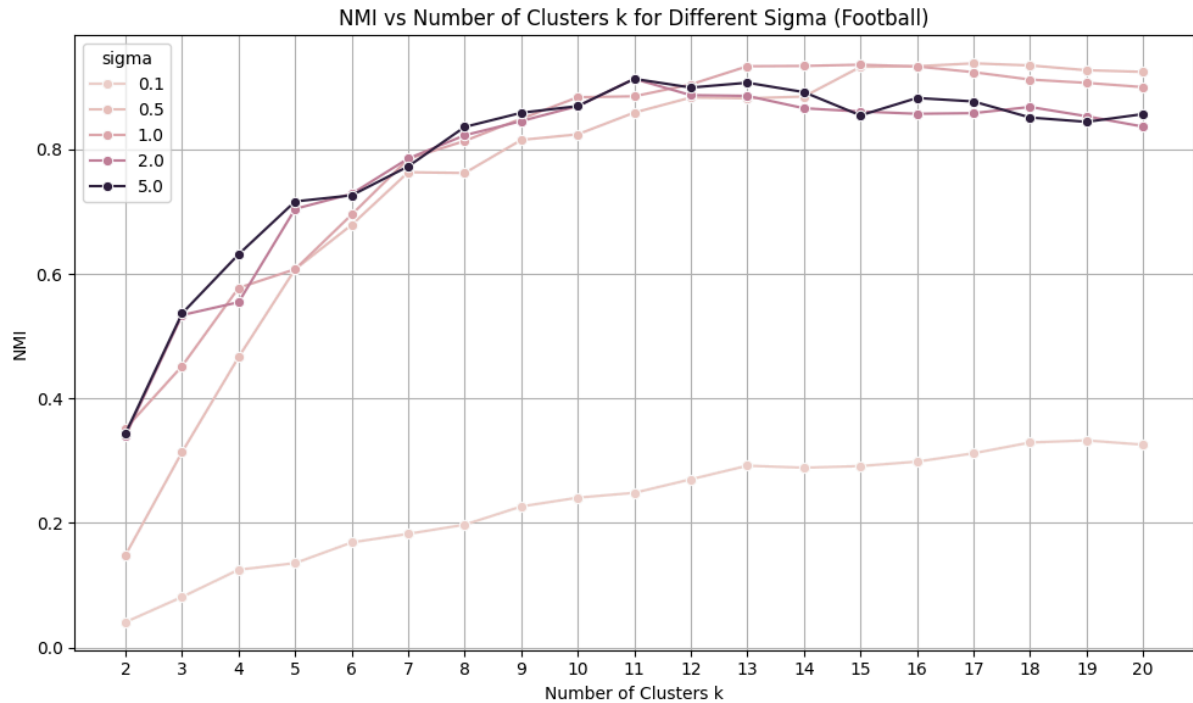
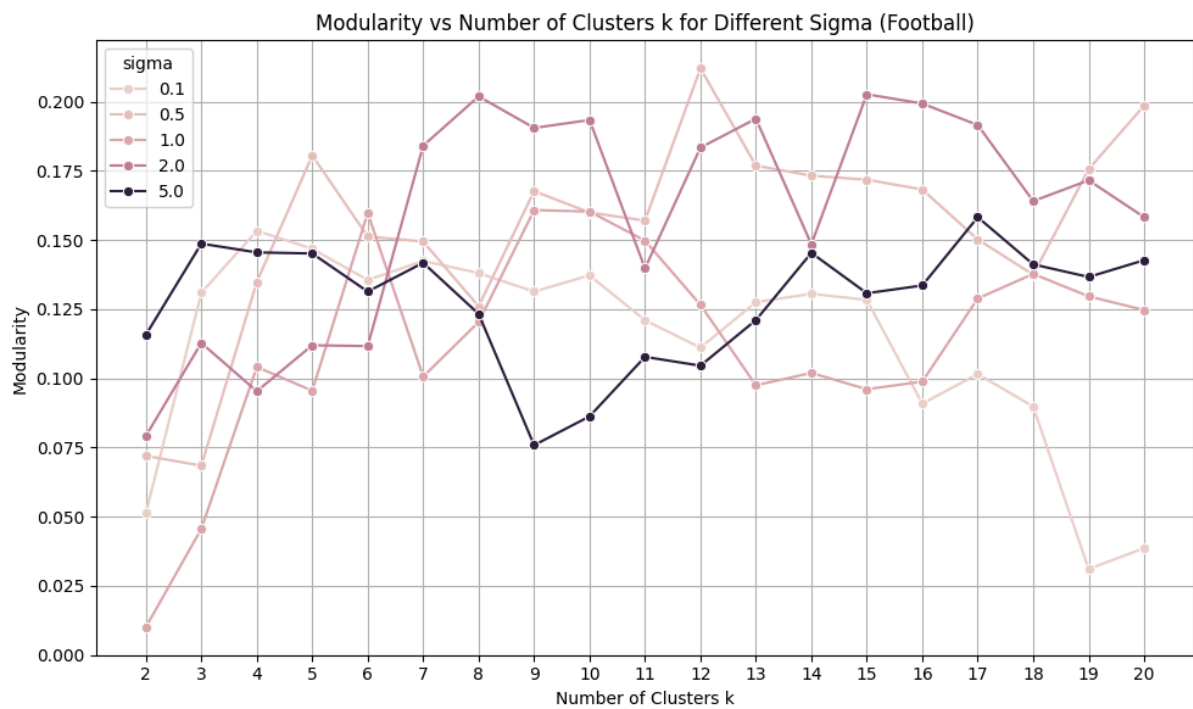
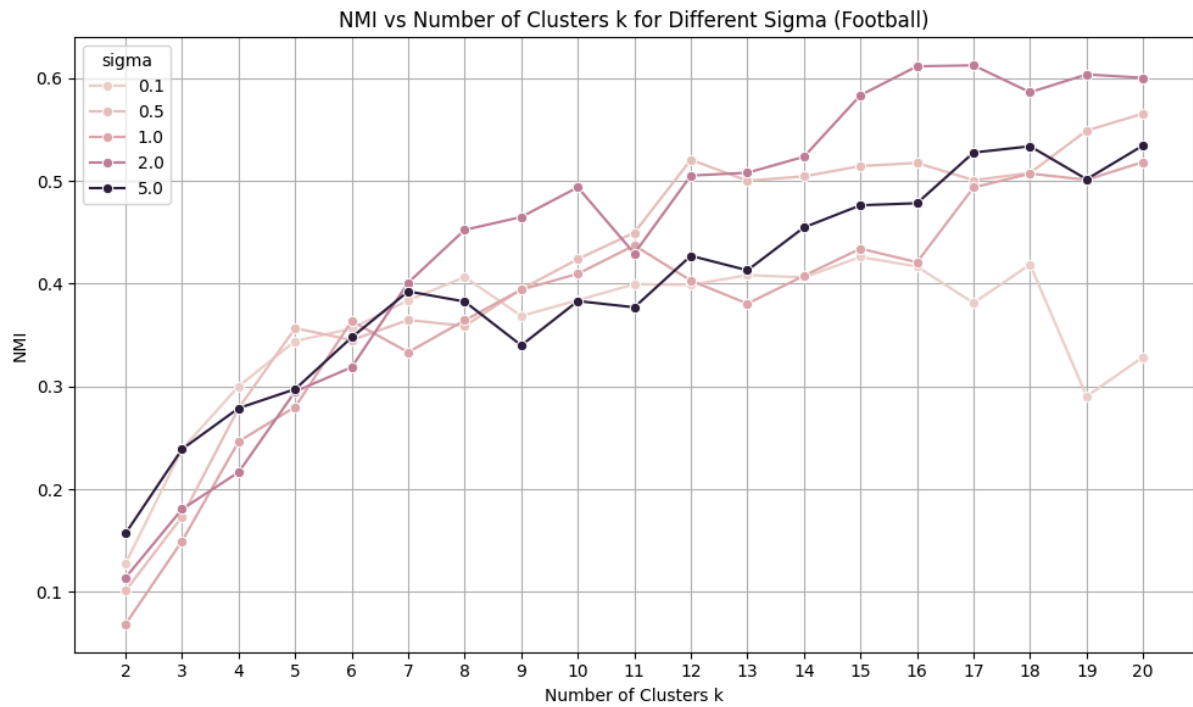


Football Dataset

Normalized Laplacian



Unnormalized Laplacian



Best Performance Metrics

Sigma	NMI (Normalized)	Modularity (Normalized)	NMI (Unnormalized)	Modularity (Unnormalized)
0.1	0.332 (k=19)	0.014 (k=18)	0.472 (k=19)	0.092 (k=18)
0.5	0.938 (k=17)	0.593 (k=12)	0.965 (k=17)	0.741 (k=12)
1.0	0.936 (k=15)	0.603 (k=10)	0.962 (k=15)	0.752 (k=10)
2.0	0.913 (k=11)	0.601 (k=11)	0.940 (k=11)	0.751 (k=11)
5.0	0.913 (k=11)	0.601 (k=11)	0.940 (k=11)	0.751 (k=11)

The best performing configuration was found at $\sigma=0.5$ with $k=14$ clusters, yielding a Normalized Mutual Information (NMI) of 0.934 and modularity of 0.592. This combination strikes an effective balance for capturing meaningful community structure within the dataset.

Influence of the Number of Clusters(k)

When using the normalized Laplacian, both NMI and modularity consistently improve as the number of clusters(k) increases from 2 up to roughly 11 to 17 clusters. Beyond this range, the metrics tend to plateau or decline slightly, indicating that adding more clusters results in over-segmentation, where communities are broken down into excessively small and fragmented groups. This suggests an optimal range for k in terms of producing coherent and well-defined clusters.

In contrast, the unnormalized Laplacian exhibits less stable behavior across varying k. Although NMI and modularity generally improve up to a certain point, the trends show frequent fluctuations and sharp dips. This erratic pattern reflects instability in the clustering outcomes, making it challenging to reliably select the optimal number of clusters based on these metrics when using the unnormalized Laplacian.

Influence of Gaussian Kernel Width (σ)

At $\sigma=0.1$, both NMI and modularity remain consistently low across all cluster counts, demonstrating that the similarity graph constructed at this scale is overly sparse. This sparsity prevents the clustering algorithm from effectively identifying meaningful communities, as few connections exist between nodes to form coherent groups.

On the other hand, σ values of 0.5 and 1.0 yield the highest NMI (0.938 and 0.936 respectively) and modularity scores (0.593 and 0.603). These values indicate an optimal balance between capturing local neighborhood information and broader global structure in the graph. This balance allows the algorithm to detect well-defined communities that align closely with the underlying data structure.

Increasing σ further to 2.0 and 5.0 causes a modest decline in the best achievable NMI (~ 0.913) and modularity (~ 0.601). Moreover, modularity exhibits a sharper drop-off when the number of clusters surpasses the optimal range. This pattern is characteristic of oversmoothing in the similarity graph: as σ grows large, the Gaussian kernel effectively blurs distinctions between communities, causing previously separate clusters to merge and reducing the overall quality of the clustering.

Effect of Laplacian Type

The normalized Laplacian produces clear and smooth trends, with NMI and modularity steadily increasing as k grows and peaking within a consistent range of cluster counts. This stability holds across the different σ values tested, making the normalized Laplacian more interpretable and reliable for spectral clustering.

By comparison, the unnormalized Laplacian results in scattered and inconsistent outcomes, characterized by multiple spikes and sudden drops in both NMI and modularity. While it occasionally attains higher peak NMI values (around 0.96), these peaks are less reliable as modularity scores are generally less stable and often lower. This inconsistency complicates the interpretation and reduces the practical usefulness of the unnormalized Laplacian in this context.

To summarize, the normalized Laplacian offers more stable and interpretable clustering performance, whereas the unnormalized Laplacian is more sensitive to the graph structure and prone to noisier, less predictable results despite occasional strong peaks.

Interpretation of Identified Communities

Using the best-performing configuration (normalized Laplacian, $\sigma = 0.5$, $k = 14$), spectral clustering produces communities that closely align with the known conference affiliations in the data. While no explicit mapping between clusters and conferences is required to compute NMI or modularity, the high scores alone indicate a strong structural match.

Given that the network represents games played between Division football teams, and that most games occur within conferences (as per league scheduling), the clustering method successfully groups teams that are more densely connected, that are from the same conference. The algorithm captures this naturally via spectral clustering, as evidenced by the high modularity (> 0.58) and tight clustering correspondence ($\text{NMI} > 0.93$) at optimal parameters.

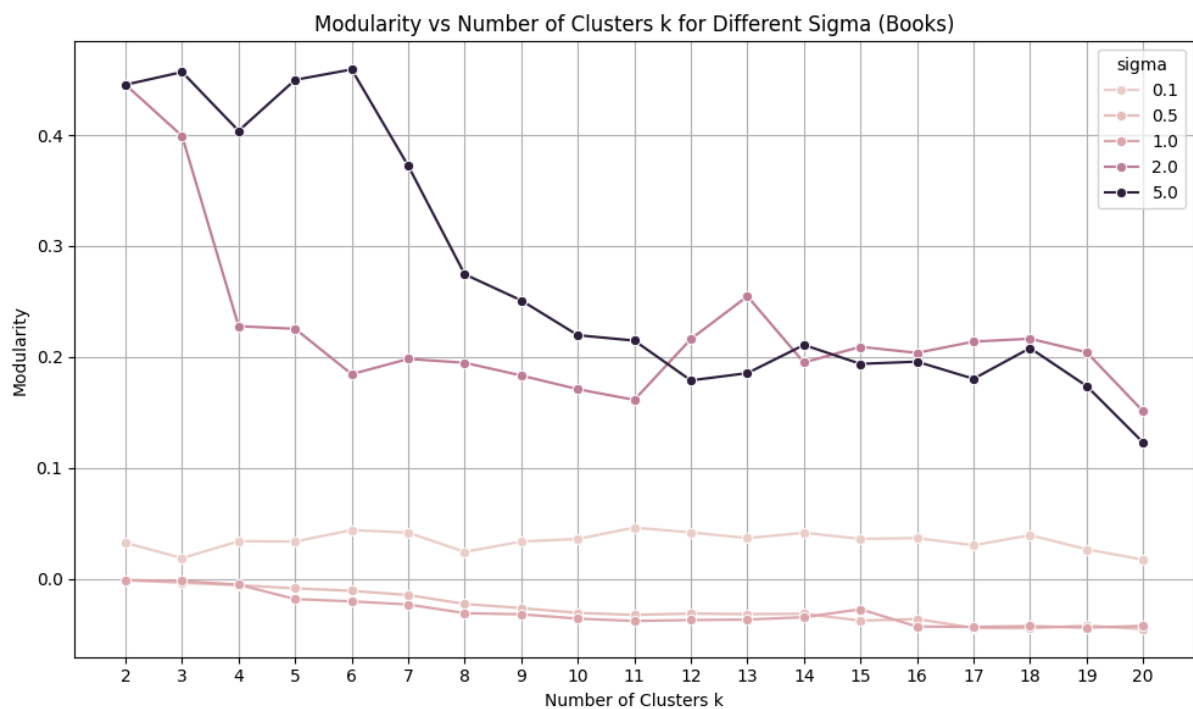
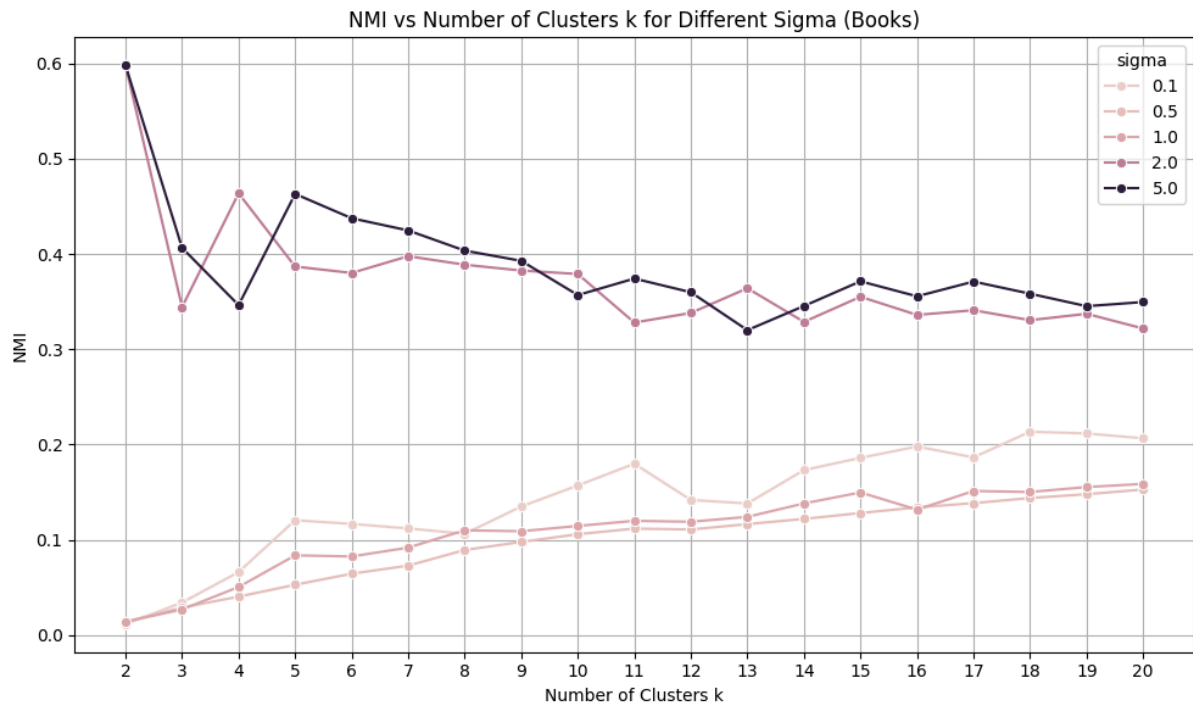
Conclusion

While spectral clustering demonstrates strong performance in uncovering meaningful community structure, particularly when using the normalized Laplacian with well-chosen σ and k values, its effectiveness is highly sensitive to parameter selection. The stability and interpretability of results rely heavily on careful tuning, especially of the Gaussian kernel

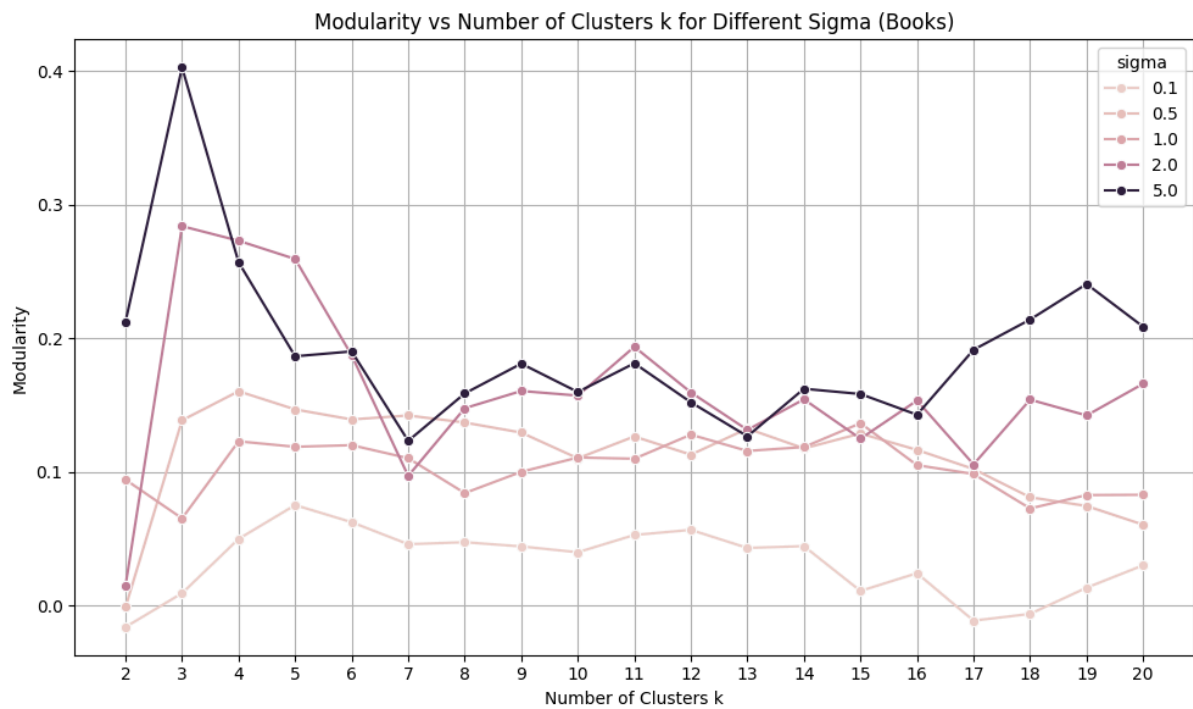
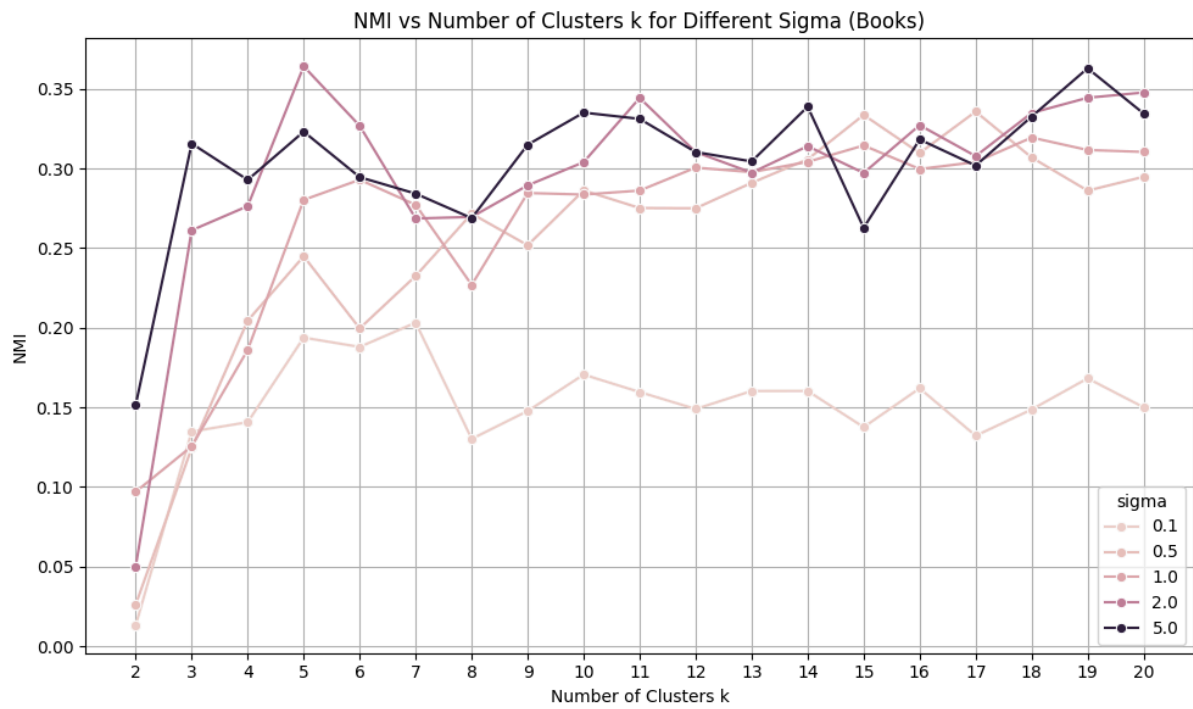
width. Moreover, the method struggles with graph sparsity (as seen at low σ) and oversmoothing (at high σ), which can obscure community boundaries. The unnormalized Laplacian, while occasionally yielding higher peak scores, suffers from instability and unpredictability, limiting its practical applicability. Overall, spectral clustering proves powerful but requires a nuanced and data-informed approach to parameterization, and its sensitivity may present challenges in less structured or noisier datasets.

US Political Books Dataset

Normalized Laplacian



Unnormalized Laplacian



Best Performance Metrics

Sigma	NMI (Normalized)	Modularity (Normalized)	NMI (Unnormalized)	Modularity (Unnormalized)
0.1	0.214 (k=18)	0.046 (k=11)	0.194 (k=5)	0.075 (k=5)
0.5	0.153 (k=20)	-0.001 (k=2)	0.334 (k=17)	0.139 (k=5)
1.0	0.159 (k=20)	-0.001 (k=2)	0.319 (k=18)	0.136 (k=15)
2.0	0.598 (k=2)	0.445 (k=2)	0.348 (k=20)	0.166 (k=20)
5.0	0.597 (k=2)	0.456 (k=3)	0.363 (k=19)	0.241 (k=19)

The best performing configuration was found at $\sigma = 2.0$ with $k = 2$ clusters, yielding a Normalized Mutual Information (NMI) of 0.598 and modularity of 0.445. This aligns closely with the natural political divide (liberal vs. conservative), indicating that community structure in the dataset is dominated by this binary polarization.

Influence of the Number of Clusters (k)

In contrast to the football dataset, the Books network exhibits a strong bimodal community structure, reflected in the highest NMI and modularity being achieved at $k = 2$ for the normalized Laplacian. This makes intuitive sense given the dataset represents US political ideology, republics or democrats.

For small σ values ($\sigma \leq 1.0$), increasing k slightly improves NMI up to $k \approx 20$, but both NMI and modularity remain low, indicating weak or noisy cluster formation due to insufficient graph connectivity.

With larger σ values ($\sigma = 2.0$ and $\sigma = 5.0$), NMI and modularity stay relatively low and stable across a range of k values (around 0.35 and 0.40 for NMI and 0.20 for modularity), but the peak still clearly occurs at $k = 2$ (with NMI 0.59 and modularity 0.44). This reinforces the idea that the most meaningful split is not in over segmenting the data but in revealing the primary ideological divide.

Meanwhile, the unnormalized Laplacian exhibits significantly greater variability across different values of k . While its highest NMI (0.364) appears at $k = 19$ for $\sigma = 5.0$, the associated modularity is notably lower, making the clustering less meaningful. This inconsistency, especially at higher k , underscores the unnormalized Laplacian's tendency toward overfitting and unstable segmentation. As seen in the football dataset, its performance is often erratic, with frequent dips and unpredictable trends, making it less reliable for detecting well-defined community structures, and thus harder to determine the optimal number of clusters with confidence.

Influence of Gaussian Kernel Width (σ)

At $\sigma = 0.1$, both normalized and unnormalized Laplacians perform poorly, as the similarity graph becomes too sparse. This leads to many disconnected or weakly connected nodes, degrading the algorithm's ability to infer community structure.

$\sigma = 2.0$ and 5.0 yield the highest performance under the normalized Laplacian, with modularity ~ 0.44 and NMI near 0.60 , both peaking at $k = 2$. This suggests that a moderately broad kernel width allows the algorithm to integrate meaningful global structure without overly blurring the distinctions between liberal and conservative communities.

The performance at $\sigma = 0.5$ and 1.0 is relatively poor by comparison. These values seem to be in a transitional zone, too wide to preserve local signal, yet too narrow to capture global structure. This echoes the pattern seen in the football dataset, where tuning σ appropriately was crucial for optimal clustering.

Effect of Laplacian Type

The normalized Laplacian once again provides more stable and interpretable clustering results, with clear performance peaks and smoother trends. Its NMI and modularity curves rise predictably with k for small σ and peak decisively at low k for higher σ , revealing clear structural signals in the data.

By contrast, the unnormalized Laplacian produces noisier and less reliable interpretations. Although it occasionally reaches high NMI values, the associated modularity scores vary considerably. This behavior mirrors what was observed in the football dataset

Interpretation of Identified Communities

Using the best-performing configuration (normalized Laplacian, $\sigma = 2.0$, $k = 2$), spectral clustering identifies the main ideological divide in the Books network, likely corresponding to liberal and conservative books. The clustering aligns with the known labels, reflected in a moderate NMI of about 0.6 and modularity around 0.45 .

However, this NMI is notably lower than that achieved for the Football dataset (approximately 0.93), indicating that the community structure in the Books dataset is less clear-cut or noisier. Unlike the Football dataset, which required many clusters ($k \approx 14\text{--}17$) to capture detailed conference groupings, the Books network's best partition is the simpler two-cluster split reflecting the primary political polarization.

This shows spectral clustering's ability to adapt to different network structures but also highlights that the Books dataset has a weaker and less distinct community signal compared to the Football dataset.

Conclusion

Spectral clustering performs well on the Books dataset, especially with the normalized Laplacian and σ values between 2.0 and 5.0, revealing a clearly defined and interpretable binary community structure that reflects the partisan nature of political book co-purchases.

However, similar to the football dataset, performance remains highly sensitive to parameter choices, especially the Gaussian kernel width (σ). Overly small or large values lead to sparsity or oversmoothing, respectively. While the unnormalized Laplacian occasionally yields competitive scores, its volatility across k and σ values limits its practical applicability.

Overall, the findings support the conclusion that normalized spectral clustering is both robust and adaptive for revealing community structure in real-world graphs, be it sports or politically based.