1. Who works together PAIRS (sponsor-cosponsor & cosponsor-cosponsor pairs)
2. Who works together GROUPS
3. What TOPICS these groups cover

**Who Works Together?**

To explore legislative collaboration, we analyzed sponsorship and co-sponsorship patterns among legislators to determine who frequently works together on bills. We adopted a three-part approach involving analyzing co-sponsorship pairs, detecting tightly connected groups using network analysis techniques, and identifying themes and keywords defining each group using NLP methods.

**(1) Analyzing Collaborative Pairs**

To quantify collaboration, we counted instances where a legislator co-sponsored a bill with another legislator in some form. Specifically, we took the frequency of both occurrences of a primary sponsor and individual co-sponsors, as well as co-sponsor-to-co-sponsor working together on a bill.

By aggregating these counts, we determined the most frequent legislator partnerships. The top 10 strongest partnerships are listed below:

| **Legislator** | **Legislator** | **Bills Worked on Together (Frequency)** |
| --- | --- | --- |
| Jason M. Lewis | James B. Eldridge | 936 |
| James B. Eldridge | Kay Khan | 823 |
| James B. Eldridge | Denise Provost | 821 |
| Jason M. Lewis | Denise Provost | 765 |
| Kay Khan | Denise Provost | 727 |
| Lindsay N. Sabadosa | Natalie M. Higgins | 719 |
| James B. Eldridge | Mary S. Keefe | 698 |
| Denise Provost | David M. Rogers | 689 |
| Marjorie C. Decker | James B. Eldridge | 683 |
| Jack Patrick Lewis | James B. Eldridge | 681 |

Notable names include James B. Eldridge and Denise Provost.

**(2) Network Analysis of Legislative Collaboration**

Using network visualization techniques, we created a graph where nodes represented legislators, and edges indicated co-sponsorship relationships weighted by frequency. For the sake of visualization, we limited our analysis to the top 100 legislators based on co-sponsorship frequency. The network graph revealed clusters of legislators who consistently co-sponsored legislation, indicating strong collaborative relationships.

*[q2 paper plot 1.png] in gdrive folder*

To further investigate collaborative tendencies, we applied the Louvain community detection algorithm. The Louvain algorithm works by iteratively optimizing the modularity of the network, grouping legislators who have the highest density of co-sponsorships. By doing so, it uncovers communities of legislators who are more strongly connected to each other than to others. These groups are likely formed based on shared policy interests, party affiliations, or committee assignments. This analysis revealed distinct communities of legislators who consistently work together, grouping legislators into clusters ranging from 3 to 22 individuals.

*[q2 paper plot 2.png] in gdrive folder*

Some of the most notable groups include:

* Group 2 (16 members): Featuring legislators such as Jason M. Lewis, Kay Khan, and Marjorie C. Decker.
* Group 1 (10 members): Including James B. Eldridge and Sal N. DiDomenico.
* Group 3 (22 members): Legislators include Lindsay N. Sabadosa and Natalie M. Higgins.

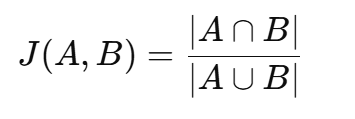
**(3) Cluster Focus Topics using NLP Techniques**

Next, our goal is to identify the key topics of different communities of legislators and analyze the policy tendencies of specific individuals. We extracted keywords from the `bill\_title` column in the provided large dataset. We calculated the most frequently mentioned keywords for each individual (both sponsors and cosponsors) and for each group identified earlier. This allowed us to construct features representing policy focus areas. During this process, we removed stop words and terms that do not carry meaningful policy implications, such as “authority” and “Massachusetts.”

At the individual level, we calculated keyword frequencies by analyzing all bills associated with each legislator and identifying their 10 most frequently mentioned keywords. It formed a personal keyword dataset, reflecting individual legislators’ policy interests. At the group level, we aggregated the keywords of all members within each group from the personal keyword dataset, calculated keyword frequencies, and identified the 10 most frequently mentioned keywords per group. The following table presents the top 10 focus topics for each group.

| **Group** | **Top 10 Keywords** |
| --- | --- |
| Group 0 | health, public, care, access, housing, commonwealth, education, safety, protect, promoting |
| Group 1 | health, public, care, access, commonwealth, education, safety, services, program, promote |
| Group 2 | health, access, public, care, commonwealth, education, services, protect, promote, safety |
| Group 3 | public, health, access, commonwealth, care, education, safety, covid, promote, program |
| Group 4 | health, public, access, commonwealth, education, care, town, promote, protect, housing |
| Group 5 | health, access, public, commonwealth, care, education, services, protect, promote, safety |
| Group 6 | health, public, care, education, access, commonwealth, services, program, children, protect |
| Group 7 | public, health, commonwealth, care, safety, education, access, school, state, program |
| Group 8 | health, public, care, access, commonwealth, education, safety, services, protect, promote |
| Group 9 | health, public, care, education, commonwealth, safety, access, program, promote, services |
| Group 10 | health, public, care, education, access, promote, commonwealth, protect, services, safety |
| Group 11 | health, public, care, commonwealth, access, education, services, safety, program, protect |
| Group 12 | health, public, access, commonwealth, care, education, services, protect, promote, safety |
| Group 13 | public, health, care, access, education, commonwealth, program, safety, services, promote |

Finally, we evaluated the alignment between individual legislators and their respective groups using Jaccard similarity, which allowed us to identify legislators whose topic interests significantly deviate from their group. Jaccard similarity is a metric used to measure the similarity between two sets, and its formula is as in:

\begin{equation}

J(A, B) = \frac{|A \cap B|}{|A \cup B|}

\end{equation}

where A represents the set of a legislator’s top 10 keywords, and B represents the set of top 10 keywords for the legislator’s group. The lower the J(A, B) value, the more the legislator’s policy focus deviates from their group's. Based on this, we filtered out legislators with a Jaccard similarity below 0.5, which indicates that these individuals have a significant difference in policy focus compared to their group. Our results show that, among the top 100 legislators based on co-sponsorship frequency, two legislators exhibited a notable discrepancy between their policy focus and that of their group. John H. Rogers from Group 8 had a Jaccard similarity of 0.43, while Todd M. Smola from Group 7 had a Jaccard similarity of only 0.33. The following table presents the topics these two legislators focus on and highlights the topics they prioritize more than their groups.

| **Legislator** | **Group** | **Top 10 Focus Topics** | **Additional Focus Topics** |
| --- | --- | --- | --- |
| John H. Rogers (Similarity = 0.43) | Group 8 | public, health, education, care, services, covid, commonwealth, emergency, program, town | covid, emergency, program, town |
| Todd M. Smola (Similarity = 0.33) | Group 7 | public, commonwealth, safety, service, services, fund, education, town, benefits, state | service, services, fund, town, benefits |

**Overall Findings**

Our analysis for this question shows distinct patterns in legislators’ community and policy focus. We summarize the results in the table below:

| **Group** | **Unique Defining Characteristic** | **Members** | **Size** |
| --- | --- | --- | --- |
| **0** | **Housing & Protection Focus** – Emphasizes *housing* and *protection* alongside public health. | Jason M. Lewis, Marjorie C. Decker, Ruth B. Balser, Jennifer E. Benson, Kenneth I. Gordon, Carolyn C. Dykema, Louis L. Kafka, Lori A. Ehrlich, Frank I. Smizik, Peter V. Kocot, Sarah K. Peake, John W. Scibak | 12 |
| **1** | **Service-Oriented Programs** – Stands out with *services* and *program* focus. | Jay D. Livingstone, Kevin G. Honan, Edward F. Coppinger | 3 |
| **2** | **Strong Public Safety & Promotion** – Unique emphasis on *protect* and *promote*. | Tricia Farley-Bouvier, Christine P. Barber, David Paul Linsky | 3 |
| **3** | **COVID-19 Concern** – The only group explicitly mentioning *COVID*. | Kay Khan, Denise Provost, Mary S. Keefe, Patricia D. Jehlen, Jonathan Hecht | 5 |
| **4** | **Town & Housing Focus** – The only group highlighting *town* alongside *housing*. | Lindsay N. Sabadosa, Natalie M. Higgins, Jack Patrick Lewis, Mike Connolly, Rebecca L. Rausch, Tommy Vitolo, Carmine Lawrence Gentile, James K. Hawkins, David Henry Argosky LeBoeuf, Carlos González, Joanne M. Comerford, Tram T. Nguyen, Brian W. Murray, Adrian C. Madaro, Maria Duaime Robinson, Michelle L. Ciccolo, Tami L. Gouveia, Mindy Domb, Peter Capano, David Allen Robertson, Natalie M. Blais, Marcos A. Devers | 22 |
| **5** | **Health & Safety Core** – General alignment but strong *safety* and *protection* presence. | Elizabeth A. Malia, Steven Ultrino, Daniel J. Ryan, Alice Hanlon Peisch, John H. Rogers | 5 |
| **6** | **Children's Welfare Priority** – Stands out with *children* in addition to *services & programs*. | Michael D. Brady, Paul W. Mark, Diana DiZoglio, Tackey Chan, Joseph W. McGonagle Jr., Frank A. Moran | 6 |
| **7** | **Education & State-Level Policies** – Highlights *school* and *state* policy concerns. | David M. Rogers, Sean Garballey | 2 |
| **8** | **Balanced & Generalist Approach** – Covers all key topics without any highly unique keywords. | James J. O'Day, Daniel M. Donahue, Antonio F. D. Cabral | 3 |
| **9** | **Program-Centric Approach** – *Program* appears uniquely high in the ranking. | Chris Walsh, Paul R. Heroux, Michelle M. DuBois, Barbara A. L'Italien, José F. Tosado, Aaron Vega, Danielle W. Gregoire, Kate Hogan, John J. Lawn Jr. | 9 |
| **10** | **Strong Promotion & Protection** – Prioritizes *promote* and *protect* together. | James B. Eldridge, Sal N. DiDomenico, Thomas M. Stanley, Michael O. Moore, Patrick M. O'Connor, Michael J. Barrett, Walter F. Timilty, Anne M. Gobi, Brendan P. Crighton | 9 |
| **11** | **Service & Safety Advocacy** – Strong focus on *services* and *safety*. | Paul McMurtry, Denise C. Garlick, James Arciero, William C. Galvin | 4 |
| **12** | **Emphasis on Protection & Promotion** – Mirrors Group 2 but broader. | Julian Cyr | 1 |
| **13** | **Program & Education-Driven** – *Program* is emphasized along with education. | Susan Williams Gifford, Steven S. Howitt, Todd M. Smola, Elizabeth A. Poirier, Bruce E. Tarr, Kimberly N. Ferguson, Brian M. Ashe, Angelo J. Puppolo Jr., Hannah Kane, Mathew J. Muratore, Angelo L. D'Emilia, Colleen M. Garry, Josh S. Cutler, Bradley H. Jones Jr., James J. Dwyer, Carole A. Fiola | 16 |

While most legislators align closely with their cooperative groups, some individuals exhibit notable deviations in their policy priorities. Especially in larger groups (such as Group 7 and Group 8), there is a tendency for certain individuals to diverge significantly from their group's primary focus. We believe that identifying these discrepancies provides insight into potential special interests and independent policy agendas. These deviations may indicate legislators pursuing issue areas outside their group due to district-specific concerns, cross-party cooperation, or external influences such as lobbying and political alliances.

| **Group** | **Unique Defining Characteristic** | **Members** | **Size** |
| --- | --- | --- | --- |
| 0 | Housing & Protection Focus | Jason M. Lewis, Marjorie C. Decker, Ruth B. Balser, Jennifer E. Benson, Kenneth I. Gordon, Carolyn C. Dykema, Louis L. Kafka, Lori A. Ehrlich, Frank I. Smizik, Peter V. Kocot, Sarah K. Peake, John W. Scibak | 12 |
| 1 | Service-Oriented Programs | Jay D. Livingstone, Kevin G. Honan, Edward F. Coppinger | 3 |
| 2 | Strong Public Safety & Promotion | Tricia Farley-Bouvier, Christine P. Barber, David Paul Linsky | 3 |
| 3 | COVID-19 Concern | Kay Khan, Denise Provost, Mary S. Keefe, Patricia D. Jehlen, Jonathan Hecht | 5 |
| 4 | Town & Housing Focus | Lindsay N. Sabadosa, Natalie M. Higgins, Jack Patrick Lewis, Mike Connolly, Rebecca L. Rausch, Tommy Vitolo, Carmine Lawrence Gentile, James K. Hawkins, David Henry Argosky LeBoeuf, Carlos González, Joanne M. Comerford, Tram T. Nguyen, Brian W. Murray, Adrian C. Madaro, Maria Duaime Robinson, Michelle L. Ciccolo, Tami L. Gouveia, Mindy Domb, Peter Capano, David Allen Robertson, Natalie M. Blais, Marcos A. Devers | 22 |
| 5 | Health & Safety Core | Elizabeth A. Malia, Steven Ultrino, Daniel J. Ryan, Alice Hanlon Peisch, John H. Rogers | 5 |
| 6 | Children's Welfare Priority | Michael D. Brady, Paul W. Mark, Diana DiZoglio, Tackey Chan, Joseph W. McGonagle Jr., Frank A. Moran | 6 |
| 7 | Education & State-Level Policies | David M. Rogers, Sean Garballey | 2 |
| 8 | Balanced & Generalist Approach | James J. O'Day, Daniel M. Donahue, Antonio F. D. Cabral | 3 |
| 9 | Program-Centric Approach | Chris Walsh, Paul R. Heroux, Michelle M. DuBois, Barbara A. L'Italien, José F. Tosado, Aaron Vega, Danielle W. Gregoire, Kate Hogan, John J. Lawn Jr. | 9 |
| 10 | Strong Promotion & Protection | James B. Eldridge, Sal N. DiDomenico, Thomas M. Stanley, Michael O. Moore, Patrick M. O'Connor, Michael J. Barrett, Walter F. Timilty, Anne M. Gobi, Brendan P. Crighton | 9 |
| 11 | Service & Safety Advocacy | Paul McMurtry, Denise C. Garlick, James Arciero, William C. Galvin | 4 |
| 12 | Emphasis on Protection & Promotion | Julian Cyr | 1 |
| 13 | Program & Education-Driven | Susan Williams Gifford, Steven S. Howitt, Todd M. Smola, Elizabeth A. Poirier, Bruce E. Tarr, Kimberly N. Ferguson, Brian M. Ashe, Angelo J. Puppolo Jr., Hannah Kane, Mathew J. Muratore, Angelo L. D'Emilia, Colleen M. Garry, Josh S. Cutler, Bradley H. Jones Jr., James J. Dwyer, Carole A. Fiola | 16 |