Advanced Cypher Queries for Social Botnet Analytics

1. Community Sizes

```
MATCH (n:Person)
RETURN n.community AS community, count(*) AS size
ORDER BY size DESC
```

Explanation: This query counts the number of nodes in each detected community (cluster). It helps to understand the relative size of each group and identify the largest or smallest communities in the social network.

2. Bots per Community

```
MATCH (n:Person)
WHERE n.is_bot = true
RETURN n.community AS community, count(*) AS num_bots
ORDER BY num_bots DESC
```

Explanation: This query finds out how many bots exist in each community. It is useful for identifying bot-dominated communities or detecting clusters where bots are most active.

3. Show All Bots and Their Communities

```
MATCH (n:Person)
WHERE n.is_bot = true
RETURN n.id, n.community
ORDER BY n.community
```

Explanation: Lists every bot node in the network, along with the community (cluster) they belong to. Useful for targeted inspection or exporting a bot list by group.

4. Most Connected (Influential) Nodes

```
MATCH (n:Person)
RETURN n.id, size((n)--()) AS degree, n.community
ORDER BY degree DESC
LIMIT 10
```

Explanation: Finds the top 10 nodes with the highest number of connections (degree centrality). These are the most "influential" or "central" actors in the social network, possibly hubs or important accounts (could be bots or humans).

5. Communities With Both Bots and Humans

```
MATCH (n:Person)
WITH n.community AS community, collect(DISTINCT n.is_bot) AS bot_types WHERE
size(bot_types) > 1 RETURN community, bot_types
```

Explanation: Finds communities containing both bots and non-bots. This can reveal clusters where bots are mixing with genuine users, which may be critical for detecting coordinated botnet influence.

6. Inter-Community Bridges

```
MATCH (a:Person)-[r]-(b:Person)
WHERE a.community <> b.community
RETURN a.id AS from, a.community AS from_comm, b.id AS to, b.community AS to_comm
LIMIT 25
```

Explanation: Finds connections (edges) that link different communities. These "bridges" can be important for information flow, rumor spreading, or botnet infiltration between groups.

7. Visualize a Specific Community

```
MATCH (n:Person)-[r]-(m:Person)
WHERE n.community = 0 AND m.community = 0
RETURN n, r, m
```

Explanation: Visualizes the full subgraph for a specific community. Useful for in-depth cluster analysis, structural exploration, or focused visualization.

8. Subgraph of Only Bots

```
MATCH (n:Person)-[r]-(m:Person)
WHERE n.is_bot = true AND m.is_bot = true
RETURN n, r, m
```

Explanation: Shows the structure of the bot-only subgraph. Helps to see if bots are highly interconnected, working in groups, or acting independently.

9. Bots vs human

```
MATCH (u:User)
RETURN
  u.is_bot    AS isBot,
  count(*)    AS count
ORDER BY u.is_bot DESC;
```

Response:

isBot	count
true	2748
false	7451

Explanation:
Bots make up about 27% of your entire network (2 748 bots vs. 7 451 humans). That's a substantial minority—enough that communities could vary widely in bot concentration.

10. Top 10 communities by bot ratio (only those with at least 20 members):

```
MATCH (u:User)
WHERE u.cluster IS NOT NULL
WITH
  u.cluster                    AS cluster,
  sum(CASE WHEN u.is_bot THEN 1 ELSE 0 END) AS botCount,
  count(*)                     AS size
WHERE size >= 20
RETURN
  cluster,
  size,
  toFloat(botCount)/size       AS botRatio,
  botCount
ORDER BY botRatio DESC
LIMIT 10;
```

Response:

cluster	size	botRatio	botCount
39	45	0.8888888888888888	40
89	82	0.7317073170731707	60
87	24	0.7083333333333334	17
38	407	0.7076167076167076	288
65	480	0.6333333333333333	304
18	225	0.6177777777777778	139
19	237	0.6118143459915611	145
71	86	0.6046511627906976	52
68	22	0.5454545454545454	12
61	106	0.49056603773584906	52

Explanation:
These results are striking:

- Cluster 39 (45 members) is almost pure-bot: 40/45 → 88.9 % bots.
- Cluster 89 (82 members) is 73.2 % bots.
- Cluster 87 (24 members) is 70.8 % bots.

Even some large clusters—like 38 (407 members, 288 bots → 70.8 %) and 65 (480 members, 304 bots → 63.3 %)—are dominated by bots.

We've uncovered several "bot-dense" communities of varying sizes. Cluster 39 looks like a tightly knit bot network of ~45 accounts; clusters 38 and 65 are massive bot-majority groups, potentially coordinating at scale.

11. Inspect Cluster 39

```
MATCH (u:User)-[:FOLLOWS]->(v:User)
WHERE u.cluster = 39
RETURN u.user_id AS user, count(v) AS outDegree
ORDER BY outDegree DESC
LIMIT 5;
```

Response:

user	outDegree
3632	3252
2875	4
353	3
1696	3
233	3

Explanation:

That first number is interesting: **User 3632** follows **3 252** accounts, even though Cluster 39 only has 45 members! The rest follow 4 or fewer.

- **User 3632** is acting like a broadcast hub: it's connected far beyond its own cluster, pushing or pulling content across the wider network.
- The other bots are almost "leaf" nodes (outDegree ≤ 4), suggesting a classic hub-and-spoke structure.

12. Target Hubs inside cluster 39

```
MATCH (u:User)-[:FOLLOWS]->(v:User)
```

```
WHERE u.cluster = 39
RETURN v.user_id     AS target,
       count(u)      AS inDegreeFromBotCluster
ORDER BY inDegreeFromBotCluster DESC
LIMIT 5;
```

Response:

target	inDegreeFromBotCluster
3632	14
6678	4
7700	4
9300	3
9843	2

Explanation:
Here we see that **user 3632** again sits at the center of this bot cluster—being followed by 14 of its 45 members—while the next most- followed targets from Cluster 39 only get 3–4 inbound links each. This reinforces the “super- hub” role of 3632: the rest of the bots overwhelmingly point at it, with very little cross- talk among the others.

13. Inspect top 5 targets

```
MATCH (u:User)
WHERE u.user_id IN [3632, 6678, 7700, 9300, 9843]
RETURN
u.user_id AS user,
u.is_bot AS isBot,
u.stance AS stance,
u.cluster AS cluster
ORDER BY user;
```

Response:

user	isBot	stance	cluster
3632	false	1	39
6678	false	1	3
7700	false	2	22

9300	false	1	41
9843	false	1	96

Explanation:
Interesting—all five of those top- targeted accounts are humans (isBot=false), and they span two stance labels (mostly stance=1, one stance=2). This tells us that Cluster 39 is overwhelmingly targeting human users, with a bias toward users holding stance 1.

14. Stance breakdown Cluster 39

```
MATCH (u:User)
WHERE u.cluster = 39
RETURN
u.stance AS stance,
count(*) AS count
ORDER BY count DESC;
```

Response:

stance	count
0	39
1	5
2	1

Explanation:

Cluster 39 is overwhelmingly composed of bots labeled with **stance 0** (39 of 45), with only a handful carrying stance 1 (5) or stance 2 (1). This suggests that:

- The bots themselves are “stance 0–aligned.”
- Yet they predominantly **target** users of stance 1 and stance 2 (as we saw earlier).

15. Top 10 communities cluster 39 is reaching out to

```
MATCH (u:User)-[:FOLLOWS]->(v:User)
WHERE u.cluster = 39 AND v.cluster <> 39
RETURN
```

```
v.cluster AS targetCluster,
count(*) AS followCount
ORDER BY followCount DESC
LIMIT 10;
```

Response:

targetCluster	followCount
86	360
23	304
57	261
22	261
41	246
70	235
91	210
65	153
66	150
96	131

Explanation:

Cluster 39's bots are overwhelmingly reaching out into these other communities:

- **Cluster 86** (360 edges), **23** (304), **57** (261), **22** (261), **41** (246)… down to **96** (131).
- The highest-contact community, **86**, is getting eight times more attention than the second-ranked cluster (360 vs. 304).

This suggests Cluster 39 either (a) views Cluster 86 as a key audience for its messaging or (b) that those members are especially vulnerable to bot outreach.

16. Profile the Top Targeted Community (Cluster 86) Size and bot rate for Cluster 86

```
MATCH (u:User)
```

```
WHERE u.cluster = 86
RETURN
count(*) AS size,
sum(toInteger(u.is_bot)) * 1.0 / count(*) AS botRatio;
```

Stance distribution in Cluster 86

```
MATCH (u:User)
WHERE u.cluster = 86
RETURN
u.stance AS stance,
count(*) AS count
ORDER BY count DESC;
```

Response:

size	botRatio
970	0.32783505154639175

stance	count
2	692
0	257
1	21

Explanation: Cluster 86 is a predominantly human community (bot ratio ≈ 32.8 %), and it's heavily "stance 2" (≈ 71 % stance 2, then 26 % stance 0, 2 % stance 1).

Bots in Cluster 39 are overwhelmingly targeting a large pool of stance-2 users. That suggests this botnet is focusing its campaigns on users aligned with stance 2.

17. How many "FOLLOWS" edges go from Cluster 86 back to Cluster 39

```
MATCH (u:User)-[:FOLLOWS]->(v:User)
WHERE u.cluster = 86 AND v.cluster = 39
RETURN count(*) AS edges86_to_39;
```

Response:

edges86_to_39
186

Explanation:
A return of **186 edges** from Cluster 86 back to Cluster 39—compared with the 360 outbound follows—means about **52 % reciprocity**. In other words, roughly half of the bot-initiated links into that big human community get followed back.

18. Which target clusters reciprocate bot-follows most?

```
MATCH (b:User)-[:FOLLOWS]->(h:User)
WHERE b.cluster = 39 AND h.cluster <> 39
WITH h.cluster AS targetCluster, count(*) AS outCount

MATCH (h2:User)-[:FOLLOWS]->(b2:User)
WHERE h2.cluster = targetCluster AND b2.cluster = 39
WITH targetCluster, outCount, count(*) AS inCount

RETURN
targetCluster,
outCount,
inCount,
toFloat(inCount)/outCount AS reciprocity
ORDER BY reciprocity DESC
LIMIT 5;
```

Response:

targetCluster	outCount	inCount	reciprocity
83	1	5	5.0
71	12	15	1.25
48	1	1	1.0
45	1	1	1.0
43	1	1	1.0

Explanation:

Those reciprocity ratios are telling:

- **Cluster 83** has only **1 outbound** follow from the bots but **5 inbound** follows—so a **500%** "return rate." That suggests a tiny, highly engaged group in Cluster 83 that not only follows bots when bots reach out, but even seeks them out unprompted.

- **Cluster 71** also returns more follows (15) than the bots send out (12), a reciprocal rate of **125%**.

- Clusters 48, 45 and 43 reciprocate exactly 100%.

This pattern points to certain communities being unusually "bot- friendly," perhaps because those users are either automated themselves or especially tuned into the botnet's content.

19. Community Cohesion

Find for each community the fraction of its edges that stay *inside* the community (a "cohesion" score):

```
MATCH (u:User)-[:FOLLOWS]->(v:User)
WITH
u.cluster AS cluster,
sum(CASE WHEN u.cluster = v.cluster THEN 1 ELSE 0 END) AS intraEdges,
count(*) AS totalEdges
RETURN
cluster,
toFloat(intraEdges) / totalEdges AS cohesionRatio,
totalEdges
ORDER BY cohesionRatio DESC
LIMIT 10;
```

Response:

cluster	cohesionRatio	totalEdges
35	0.7733691821765121	38309
36	0.6209185586145729	109367
22	0.5620634629973579	38227
23	0.5516415721126882	139074
4	0.5373560730672668	11551
86	0.436764517369568	39005
38	0.37375103625812645	22919

66	0.3719483840967217	68816
19	0.34779050736497547	12220
57	0.3397736007267137	57244

Explanation:

- Cluster 35 (≈ 77 % intra-community edges) is by far the most "insular"—over three-quarters of its follows stay within the same cluster, signaling a tight-knit echo chamber.

- Clusters 36, 22, 23 and 4 also have more than half their edges internal, indicating moderately cohesive communities.
- From Cluster 86 onward, internal cohesion drops below 50 %, meaning these groups broadcast outward more than they talk among themselves.
- In particular, Cluster 38 (0.374) and **57** (0.340) are very outward-oriented, suggesting they act as broadcasters or outreach hubs rather than closed clusters.

This hierarchy—from highly insulated (Cluster 35) to highly outward-facing (Cluster 57)—helps us classify which communities function as echo chambers versus which serve as "spokes" pushing content into the broader network.

20. Bridge-Nodes (High Inter-Cluster Degree)

Find the top 10 users who follow outside their own community most heavily:

```
MATCH (u:User)-[:FOLLOWS]->(v:User)
WITH
u.user_id AS user,
u.cluster AS cluster,
sum(CASE WHEN u.cluster <> v.cluster THEN 1 ELSE 0 END) AS interDeg
RETURN user, cluster, interDeg
ORDER BY interDeg DESC
LIMIT 10;
```

Response:

user	cluster	interDeg
3632	39	3222
1914	89	2132
2180	31	2043
311	41	2036

960	16	1866
328	43	1607
106	65	1469
643	70	1467
1494	89	1459
2893	53	1393

Explanation:
- User 3632 (Cluster 39) again tops the list, following 3222 users outside its own cluster—confirming its role as a major outreach hub.
- Users 1914 (Cluster 89), 2180 (31), 311 (41) and others each follow over 2 000 outsiders, marking them as super-bridges who connect their communities to the rest of the graph.
- The presence of multiple users from Cluster 89 (1914 and 1494) suggests that this community has several active exporters of content.
- These “bridge” nodes are prime candidates for further investigation—they likely serve as the principal conduits for information flow (or bot-driven campaigns) between clusters.