Analyzing the Venture Capital Ecosystem: a network science approach

Introduction:

The venture capital ecosystem can be represented as a network of investors and startups. We examine five major VC firms: Accel, Andreessen Horowitz, Bessemer Venture Partners, Lightspeed Venture Partners, and Sequoia Capital, and the startups they have invested in, using tools from our course. By constructing an investor–startup bipartite network (VCs connected to their portfolio companies) and then a startup–startup projection (startups connected if they share a common VC investor), we can analyze the structure of this ecosystem. We examine key network properties such as connectivity, modularity, degree distribution, and path length and discuss how these relate to the success outcomes of startups. We also dedicate a section to exploring whether the network show signs of an “AI funding bubble” in 2025 (similar to the late 1990s dot-com bubble) by looking at the clustering of investments in artificial intelligence startups and the emergence of highly central “hub” companies. We connect our findings to concepts from the course, including preferential attachment, scale-free networks, and network robustness.

Data and Network Construction:

The dataset contains 249 records of investments made by these five VC companies so far in 2025. We decided to only look at investments from 2025 and these five VC firms because of some data limitations we had. VC investment data is extremely valuable, so it is difficult to find an open source collection of it. We ended up having to manually go through records in Pitchbook and rewrite them in a CSV file. We decided to limit our manual efforts to just the year 2025, as we did not have the capacity to do much more, and we thought this year would represent this "AI bubble" we were trying to capture well. We thought that these five VC firms would be effective as well, as they are the major headliner firms in the US, which led us to believe they may have some interesting co-investment patterns. We also wanted to focus on these big-name firms that are in Sillicon Valley (all five of the VCs we chose are headquartered there), as we thought we may see some interesting co-investment patterns. In total, these firms have invested in 222 unique startups. The network is bipartite: one set of nodes represents the five VCs, and the other represents the startups. We first constructed this bipartite graph of VC–startup ties. An edge between VC X and startup Y indicates X invested in Y. In this network, each VC node is connected to all startups in its portfolio, and each startup node is connected to all its investors (some startups have multiple of the five VCs as investors). From the bipartite graph, we derive a one-mode projection connecting startups to each other. In this startup–startup network, an edge between two startups means they share at least one common investor among the five VCs. This projected network allows us to look at how startups cluster together based on investors and to identify “communities” of startups that share investors.

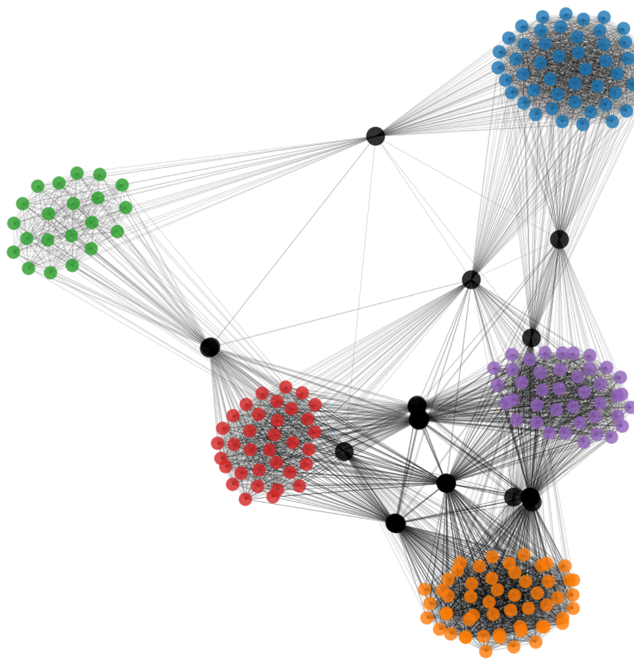
Network Size and Degree:

The bipartite network has 227 nodes (five VCs and 222 startups) and 249 VC–startup edges. Each startup’s degree in this bipartite graph is the number of these five VCs that have funded it. Most startups in the data are backed by exactly one of the five VCs, but a minority have multiple investors from this group. 199 startups (89.6%) have a single one of these VCs as an investor, 19 startups have two of them, and four startups have three. The average number of VC backers per startup is about 1.12. From the investor side, each VC’s degree is the number of startups it invested in (within this dataset from the year 2025). A16Z has the most investments in 2025 (67 startups), followed by Sequoia (55), Accel (54), Lightspeed (48), and Bessemer (25).

Building the startup–startup graph:

In the projected startup network, two startups share an edge for each common VC. Because each VC’s portfolio becomes a clique (all companies funded by the same VC become inter-connected in the projection), the structure tends to form clusters corresponding to each VC’s portfolio. If a startup has only one VC, it connects to all other startups funded by that same VC. If a startup has two (or more) of the five VCs as investors, it will lie at the intersection of those two clusters, connecting their portfolios. For example, if startup A is funded by both Sequoia and A16Z, A will have links to every startup in Sequoia’s portfolio and every startup in A16Z’s portfolio (in the projection graph). Such a startup serves as a hub or bridge between the Sequoia-funded cluster and the A16Z-funded cluster. The result is a network of startups that is highly clustered by investor, but with critical bridges connections created by co-invested companies.

Network Structure and Key Properties:



The projected startup-startup network is shown above. We observe five tight clusters of nodes corresponding to the portfolios of Accel (blue), A16Z (orange), Bessemer (green), Lightspeed (red), and Sequoia (purple). Startups funded by only one of these VCs tend to connect exclusively to others in that VC's cluster. Meanwhile, the multi-backed startups (shown as larger black nodes) sit at the intersection of clusters, linking, for example, the orange and purple groups (A16Z and Sequoia co-investments), or the purple, orange, and red groups (startups with all three of Sequoia, A16Z, and Lightspeed as investors). These bridging startups create a network that is not completely separated by VC – there is a single connected component across all 222 startups, rather than five disconnected components – but the connectivity is highly structured and modular.

Modularity:

We quantify this intuitive cluster structure using the modularity metric from community detection. The startup–startup network shows a high modularity of 0.68, when partitioned by investor portfolios. This confirms that the network naturally decomposes into communities that closely align with each VC's set of companies. A modularity near 0.7 is quite high, indicating clear portfolio clusters with relatively sparse connections between them. In practical terms, each VC has an identifiable "sphere" of influence in the startup network. The fact that modularity is not even higher (approaching 1) is due to the presence of the co-portfolio investments: whenever two VCs co-invest in a startup, they introduce links between their previously separate clusters, reducing the modularity by bringing the clusters together

Path Length and Small-World Properties:

Despite the strong clustering, the network of startups is fairly well-connected thanks to the bridges. The average path length between startups is approximately 3.50 steps. In other words, on average a startup can reach another startup through about 3 to 4 intermediary connections (shared-investor relationships). This is reminiscent of a “small-world” effect: even though most startups only share an investor with a few others, the presence of highly connected hubs (startups with multiple investors, or simply being part of a large VC portfolio) means there are short routes across the graph. In the extreme case, if a startup has three VCs invested in it, it provides a direct link among three clusters at once. The relatively short average path length (compared to a random network of similar size and degree) indicates that while the network is segmented by investor, it does not take many hops to traverse those segments. This is largely because we only are looking at five investors with a relative small number of their portfolio companies. Influence that some investors have (e.g. investor knowledge, industry connections, etc.) could thus spread between portfolio companies of different VCs via these co-investment links. For example, since the startup ElevenLabs recieved investments from Sequoia and A16Z this year, they could spread investor information between these two firms. Typically, when VC firms invest in startups, they are pretty involved in their operations, and provide guidance and mentorship to the founders. So, as the founders of ElevenLabs are receiving insights and guidance from both of these competitors, they may be able to share valuable information between the companies. Overall, it is interesting how all of these top VC firms are competing with each other, but network analysis tells us that they often invest in the same companies, so perhaps a lot of knowledge or information is being spread across the firms via the startups.

Degree Distribution and Hubs:

The degree of a node in the startup–startup network corresponds to how many other startups it is connected to via shared investors. This ranges from a minimum of 24 (for a startup in Bessemer’s portfolio, since Bessemer had 25 companies in the dataset, each Bessemer-only startup links to the other 24) up to as high as 149. The maximum degree nodes are those with three VC investors (the black hubs connecting three clusters). For instance, Wiz, Safe Superintelligence, and Saronic Technologies are each co-funded by A16Z, Lightspeed, and Sequoia – consequently, each connects to the union of all startups in those three VC portfolios, reaching 149 other startups in total. By contrast, a startup funded only by A16Z (the largest single portfolio) would connect to at most 66 others (since A16Z backed 67 startups, forming a clique of size 67 in the projection, giving each A16Z-only startup degree 66). Thus, multi-VC-backed startups serve as the high-degree hubs. This network shows a heavy-tailed degree distribution: most startups have degree on the order of tens (only one investor’s portfolio), but a few have degree over a hundred. In larger venture networks, such skewed distributions often follow a power-law form characteristic of scale-free networks, arising from preferential attachment dynamics. Preferential attachment means that new investment relationships tend to form attaching to startups that already have significant investor attention or “popularity.” In our context, if a startup gains one top VC, it likely attracts interest from other top VCs – the “rich get richer” phenomenon – leading to a few startups accumulating disproportionately many famous investors. The result is a network with hubs (a few very connected nodes) and many nodes with low degree, which is exactly what we observe. (Because our data is restricted to five investors, the maximum degree is capped by the union of five portfolios; however, if we looked at the broader VC network, we might expect a scale-free structure with some startups attracting dozens of VC investors over time)

Clustering and Connectivity:

As noted, each VC’s portfolio forms a clique in the startup projection. This means the local clustering coefficient for nodes is quite high – if startup X and Y both share investor V, and Y and Z share investor V as well, then X, Y, Z all connect to each other. Within each investor cluster, the clustering coefficient is 1 (complete graph). However, across clusters, clustering drops off. A startup that has two investors creates many connections, but two startups from different clusters might not share a second investor to form a triangle. This structure has implications for network robustness, which we discuss in a later section on AI bubble dynamics. In summary, the network structure is defined by strong community boundaries with sparse but crucial inter-community links. Each major VC acts almost like its own “hub” in the bipartite sense, but in the one-mode startup graph the hubs are the co-invested startups. The modularity of ~0.68 quantitatively confirms that the five VC portfolios are the primary divisions in the network. The presence of multi-VC backed companies lowers modularity and average path length, making the network more connected (average distance ~3.5) than it would be if VCs never co-invested. This small-world, hub-and-spoke character is common in investment networks, where a few deals (and a few companies) link otherwise separate investor communities.

Startup Success and Investor Backing:

A central question is whether a startup’s position in this network – particularly how many VC backers it has – correlates with its success. The dataset includes a “success probability” metric for each startup (an estimate of the startup’s chance of an IPO or acquisition). We compare startups funded by a single top VC versus those co-funded by multiple top VCs. The results show a clear trend: startups with multiple big-name VCs tend to have higher success probabilities on average. Startups backed by only one of the five firms have an average success probability of around 84.8%, whereas those backed by two or more of these firms average about 88.2% success probability. In other words, multi-VC-backed companies are, on average, rated ~3–4 percentage points more likely to succeed (88% vs 85%). This aligns with the intuition that when multiple prominent investors invest in the same startup, it is often a signal of quality or momentum. We cannot say for certain if there is anything causal in this regard. It could be that the firms themselves likely pick stronger startups, that would have had a higher success probability regardless. On the other hand, it also could be that many firms investing in a startup cause it to have a higher success probability by contributing resources and networks that boost the startup's odds. Notably, many of the startups with the highest success probabilities in the data are those with heavyweight syndicates. For instance, OpenAI (funded by A16Z and Sequoia) and Anthropic (funded by at least two of the five, including Lightspeed) each have a listed success probability of 98%. These companies are among the hottest AI startups and attracted multiple major investors, coinciding with confidence in their success. By contrast, startups with lower success probabilities (e.g., some below 70%) in the dataset often appear with only a single backer, perhaps reflecting higher uncertainty or earlier-stage ventures.

Signs of an AI Funding Bubble?

A fascinating aspect of the 2025 VC landscape is the prominence of AI startups. Many of the companies in our analysis, especially those with multiple major investors, are in AI or related high-tech sectors. Out of the 23 startups in our data that have more than one of the five VCs as backers, 14 (≈61%) are AI/ML-focused companies (by their described verticals). Also, if we sum the total funding raised by startups in this dataset, a disproportionate share – roughly 78% of all capital has gone into companies tagged with "Artificial Intelligence & Machine Learning". This mirrors what was happening broadly in the market: by Q1 2025, an overwhelming majority of venture funding value was flowing into AI deals (which includes VC firms outside of the five we looked at). This concentration of investments in one sector raised widespread discussion of an "AI bubble," reminiscent of the late 90s dot-com bubble when internet startups got a ton of venture capital investments.

Network Perspective on the AI Cluster:

In our startup network, the AI companies are often the bridging nodes connecting multiple investors, which implies they form a kind of central cluster. This means AI startups in 2025 weren't isolated; they were the ones linking different VC portfolios together. It seems like top VCs all chased the same AI deals, leading to co-investment ties between VCs that make the overall network more tightly knit at the top. The network thus reveals an "AI hub" phenomenon: a few companies (like OpenAI, Anthropic, Wiz, etc.) sit at short distance to most other companies through shared investors. This centrality can be quantified by measures like betweenness centrality. Indeed, the startups with highest betweenness in our data are those multi-VC AI firms (for example, Bridgetown Research and Graphite rank at the top by betweenness centrality, indicating they lie on many shortest paths connecting different portfolio clusters). The fact that 5 out of 5 top VCs were all investing in the same theme (AI) and often in the same companies suggests a degree of herding behavior. In classic bubble dynamics, investors fear missing out on the "next big thing," so they pile into whatever sector is attractive at the time. This drives valuations super high for a few companies.

Conclusion:

Network analysis provides a powerful lens on the social dynamics of venture capital. It helps us see not just individual investments, but the structure of the flow of capital and information. The framework of preferential attachment explains how we arrived at a hub-dominated network, and the notion of scale-free robustness warns us what could happen if those hubs fail. Going forward, combining network metrics with financial metrics could improve our ability to detect and perhaps even predict bubbles. For example, when we see a network becoming rapidly more central and dominated by a single community, it might be good to question the sustainability of that. Ultimately, the venture ecosystem’s goal is to balance competition and connectivity – a highly modular network (everyone sticking to their own investments and not co-investing) can stifle innovation, whereas a highly interconnected network (everyone chasing the same deals) can lead to boom-bust cycles. Our project analyzed both the benefits of connectivity (co-investments boosting success probability) and the risks of too much concentration (potential bubble). By continuing to monitor these network patterns, we can gain deeper insight into how social dynamics and network effects drive the venture capital landscape.