Chapter 2

Formulating Hypotheses



It all starts with a question...

What is an Hypothesis?

A hypothesis is a proposed explanation for a phenomenon.

It is a statement that can be tested by scientific research.

Not only making a prediction
but also identifying how to test it.



Why to we need Hypotheses?

Hypotheses provide direction to research.
They help us to:

- focus the study and
- determine the methods and analyses.

Hypotheses are essential for advancing knowledge by enabling testing and refining theories.



Examples

Psychology

Individuals who practice meditation have lower stress levels than those who do not.

Medicine

The new drug lowers blood pressure

Environmental Science

Urban areas have poorer air quality than rural areas.

Online Social Network Analysis

The presence of echo chambers in online social networks decreases the diversity of information exposure among users

Teams operating in dense social networks will generate more innovative ideas compared to teams in sparse networks.

As the size of an individual's social network increases, their likelihood of experiencing information overload also increases.

In online communities, individuals who engage in reciprocal interactions are more likely to be trusted by others than those who do not.

Exposure to negative interactions, such as cyberbullying, on social media leads to higher levels of anxiety in young adults

Examples
(non-hypotheses)

"Social media is bad for people."
Lacks specificity regarding what aspect of social media and who is affected.

"All students who study hard do well in life."
Too broad and undefined, lacking measurable criteria like what "doing well in life" means.

People who are good at math are better at solving math problems."
Circular reasoning not offering any new insight/explanation.

"There is life on another planet, but we can never detect it."
Cannot be tested or proven false, making it unscientific.

"Chocolate is the best dessert."
Subjective and based on personal opinion, making it untestable.

"If it rains, then the stock market will go up."
There's no logical or theoretical connection between the two variables.

Characteristics of a Good Hypothesis

An Hypothesis must be:

- **Testable**:
  A hypothesis must be testable through experiments or observations.
- **Falsifiable**:
  It must be possible to prove the hypothesis wrong.
- **Clear** and **Precise**:
  The hypothesis should be unambiguous and clearly defined.
- **Specific**:
  A good hypothesis is specific and focused, not broad and general.
- **Based** on **Existing Knowledge**:
  It should be grounded in current understanding and previous research.

Testable

Users with a higher number of followers are more influential in spreading information within the network.

Rationale:
Testable by: (i) analyzing follower counts and (ii) the reach/impact of posts made by users (e.g., the number of shares, likes, and comments).

Method:

- Collect data on user follower counts;
- Track the spread of specific posts;
- Determine if there's a correlation between follower count and information spread.



The overall sentiment of a user's posts can predict their future real-life events (e.g., job promotions, relationship changes)

Rationale:
This hypothesis assumes a direct causation between online sentiment and future life events.

Issues:

- External factors:
  Real-life events are influenced by a wide range of factors not captured by online data.
- Prediction complexity:
  Accurately predicting specific future events based on sentiment is extremely challenging and requires data not typically available in social networks.



Falsifiable

Users who frequently share political content on social media are more likely to participate in offline political activities (e.g., attending rallies, voting)

Rationale:
This hypothesis can be tested by surveying users about their offline political activities (e.g., from surveys) and analyzing their social media activity.

Falsifiability:
If there is no significant relationship between sharing political content and offline political participation, the hypothesis would be disproven.



Users who often post inspirational quotes are generally more spiritually fulfilled

Rationale:
While it might seem plausible that the tone of social media posts could relate to mental health, long-term mental health outcomes are influenced by factors beyond social media activity.

Non falsifiability:
The complexity and multitude of influencing factors on mental health make it impossible to design a conclusive empirical test for this hypothesis



Clear and Precise

Users who are very active on social media are more popular.

Rationale:
It is vague and imprecise because it does not clearly define "very active" or "popular." The lack of operational definitions makes it difficult to measure and test the hypothesis.

Clarity issues:
What constitutes being "very active"?
How is "popularity" measured?

Ambiguity issues:
The hypothesis does not provide specific definitions, criteria or thresholds, making it open to subjective interpretation



Users who post more than five times a day on social media have a higher engagement rate (likes, shares, comments) compared to users who post less frequently.

Rationale:
It specifies the variables being measured (posting frequency and engagement rate) and defines the threshold for frequent posting (more than five times a day)



Specific

Twitter users who include hashtags in their tweets receive more retweets compared to those who do not include hashtags.

Rationale:
It is specific because it clearly defines the variables involved (use of hashtags and number of retweets) and the expected relationship between them.



Social media activity influences user popularity.

Rationale:
It is not specific because it does not clearly define the key terms ("social media activity" and "user popularity") or the nature of the influence.

Clarity issues:
Social Media Activity could refer to a variety of actions
User Popularity could be measured in a variety of ways

Ambiguity issues:
The broad and vague nature of the hypothesis makes it difficult to draw meaningful and actionable conclusions.



Existing Knowledge

Users with strong ties (frequent interactions) on Facebook are more likely to share each other's content compared to users with weak ties (infrequent interactions).

Rationale:
It is grounded in existing knowledge from social network theory, particularly Granovetter's "strength of weak ties" theory.

Supporting Evidence:
- Research shows that strong ties, characterized by frequent interactions and high levels of trust, are more likely to engage in reciprocal behaviors such as sharing content.
- Previous studies have demonstrated that users who frequently interact on social media are more likely to engage with each other's posts.



Users who like the color blue are more likely to have more followers on Instagram.

Rationale:
It is not based on existing knowledge or theories related to social network analysis or user behavior

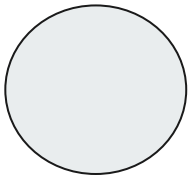
Issues:
- How to accurately determine a user's color preference is unclear
- There is no existing theory or empirical research suggesting a connection between color preference and social media popularity.



Types of Hypotheses

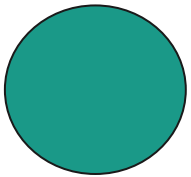
Null Hypothesis (H0)

The null hypothesis states that there is no effect or no difference.
It is the hypothesis that researchers typically aim to test against.



Alternative Hypothesis (H1)

The alternative hypothesis states that there is an effect or a difference. It is what researchers hope to support.



Directional vs. Non-Directional

Directional Hypotheses predict the direction of the effect or relationship.

Non-Directional Hypotheses predict an effect or relationship but do not specify the direction.

Examples
(Ho-H1)

Psychology

- **H1**: Individuals who practice meditation have lower stress levels than those who do not.
- **H0**: There is no difference in stress levels between individuals who practice meditation and those who do not.

Environmental Science

- H1: Urban areas have poorer air quality than rural areas.
- H0: There is no difference in air quality between urban and rural areas.

Medicine

- H1: The new drug lowers blood pressure
- H0: There is no effect of the new drug on blood pressure.

Online Social Network Analysis (exercise on the other examples)

- H1: The presence of echo chambers decreases the diversity of information exposure.
- H0: The presence of echo chambers does not affect the diversity of information exposure.

The importance of H0

1. It provides a baseline for Comparison
2. Allows to better frame H1
3. Support to Falsifiability and Testability of H1
4. Guiding Research Design and Analysis
5. It contributes to Reproducibility

Default position that there is no relationship or no effect between variables being studied.

It is structured to be falsifiable, meaning that it can be proven false if the data consistently show a pattern or effect that is unlikely to occur by chance.

H1 contrasts with H0 and represents what the researcher believes to be true or expects to find based on theory, previous research, or preliminary evidence.

H0 helps in designing studies by influencing decisions about sample size, data collection methods, and statistical analyses.

Researchers can replicate the study or build upon it, using the same or similar H0 and H1

Types of Hypotheses

Null Hypothesis (H0)

The null hypothesis states that there is no effect or no difference.
It is the hypothesis that researchers typically aim to test against.

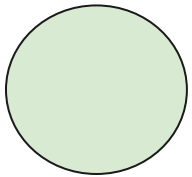
Alternative Hypothesis (H1)

The alternative hypothesis states that there is an effect or a difference. It is what researchers hope to support.

Directional vs. Non-Directional

Directional Hypotheses predict the direction of the effect or relationship.

Non-Directional Hypotheses predict an effect or relationship but do not specify the direction.



Examples
(Directional vs. Non Directional)

H1: *Increasing the frequency of social media advertising will lead to a higher number of website visits.*

Directional: This hypothesis predicts a specific direction of effect—increasing the frequency of social media advertising is expected to result in more website visits

There is existing theoretical or empirical evidence suggesting the direction of the effect.

H1: *There is a relationship between socioeconomic status and academic achievement.*

Non-Directional: This hypothesis does not specify the nature of the relationship—whether higher socioeconomic status leads to better academic achievement or vice versa.

There is ambiguity or insufficient prior knowledge about the relationship between variables.

Directional or not?

Directional Hypotheses

Focus and Specificity: provide clear guidance on the expected outcome of a study.

Theory Testing: useful when existing theories or prior research suggest a specific causal relationship or directional effect.

Hypothesis Testing: Statistical analyses used to test directional hypotheses, (e.g. one-tailed tests) are more powerful than tests for non-directional ones.

Practical Applications: In applied research directional hypotheses help in predicting outcomes of interventions or treatments.

Non Directional Hypotheses

Exploratory Research: valuable in exploratory research when there is insufficient prior knowledge about the relationship between variables.

Complex Relationships: acknowledge the possibility of unexpected relationships that might not conform to a straightforward directional prediction.

Avoiding Biases: can help avoid confirmation biases, e.g., data interpretation that fit preconceived directional expectations.

Open-Ended Research: encourage open-ended exploration potentially leading to insights or hypotheses to be further investigated with directional hypotheses.

How to Formulate a Hypothesis



Identify a Research Question

Start with a clear research question that you want to answer. Make an informed prediction. Refine the process until needed.

Conduct a Literature Review

Review existing literature to understand what is already known about the topic.

Conceptualize the Variables

Identify the independent and dependent variables in your study.

Construct the Hypothesis statement

Formulate a clear, concise, and specific hypothesis statement. Refine it until is well formed.

Common Pitfalls and How to Avoid Them

Avoiding Vagueness	Ensure the hypothesis is specific and clearly defined.
Ensuring Feasibility	The hypothesis should be testable within the constraints of your resources and timeframe.
Avoid Over Complication	Keep the hypothesis simple and focused. Avoid unnecessary complexity.

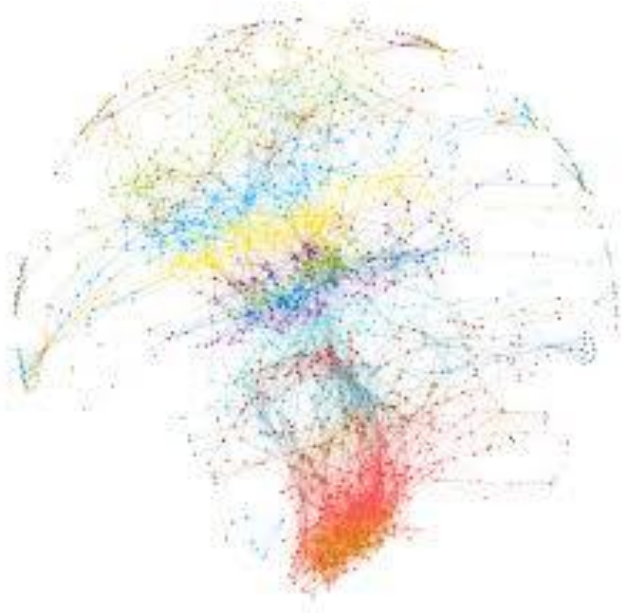
It all starts with a question...

What are you asking for?

Network Science and Hypotheses formulation

Applied Network Science studies add some **implicit assumptions** to the explicit hypotheses to be made:

- Graph modeling makes sense
- The selected graph model make sense
- The applied data transformation do not alter bias the analysis outcome (or the bias is known)



Research Objectives

Define Purpose: Clearly articulate what you aim to discover or understand through your study. This helps in forming focused and relevant hypotheses.

Specific Questions: Break down the broad research objectives into specific research questions that your hypotheses will address.

Data Availability and Quality

Data Sources: Ensure that you have access to reliable and relevant data to test your hypotheses.

Data Quality: Assess the quality and completeness of the data, considering potential biases and limitations.

Testability

Measurability: Formulate hypotheses that can be tested empirically using available data and methods.

Operational Definitions: Clearly define the variables and metrics you will use to test your hypotheses.

Complexity and Scope

Manageable Scope: Ensure the hypotheses are manageable within the scope of your resources, time, and data availability.

Balance: Strike a balance between complexity and feasibility, avoiding overly broad or overly narrow hypotheses.

Network Characteristics

Network Type: Specify the type of network (e.g., social network, biological network, technological network) and its relevant characteristics (e.g., size, density, directed/undirected).

Dynamic vs. Static: Consider whether your study will address static networks or dynamic changes over time.

Ethical Considerations

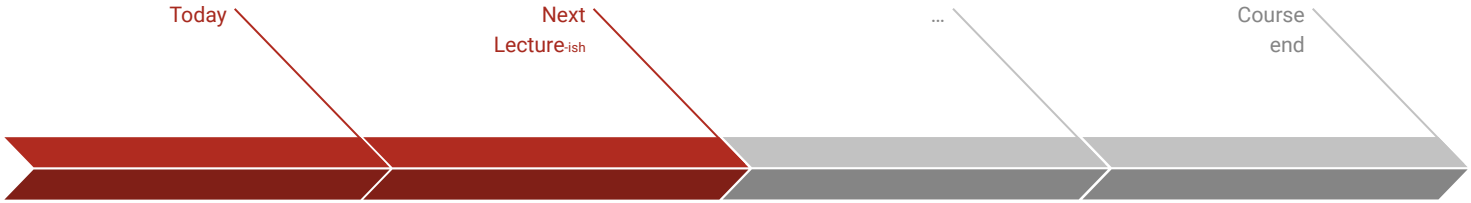
Privacy and Consent: Ensure that your study complies with ethical standards, especially when dealing with sensitive data, such as personal information in social networks.

Transparency: Be transparent about your methods and potential conflicts of interest.

What should I control for while shaping Network Study?

What’s next?

Course Project



Create a team
& Familiarize with HP formulation

Up to 3 students
Mixing competences is a plus

Select a project
& work on HP formulation

Projects/Data will be proposed by "Tutors"

Tutor will support the group that selected his/her project

Follow lectures
& work on the project

Each lecture will provide tools/techniques to frame the project, validate its result.

Laboratories will complement with hands-on sessions

Discuss your work
& live happily ever after

Chapter 2

Conclusion

Take Away Messages

1. Hypotheses have to satisfy specific characteristics
2. Without a clear hypothesis we cannot make conclusive observations

Suggested Readings

- Doing Research: A New Researcher's Guide (Ch2-3)
- https://www.enago.com/academy/how-to-develop-a-good-research-hypothesis/

What's Next

Chapter 3: Modeling Choices

