

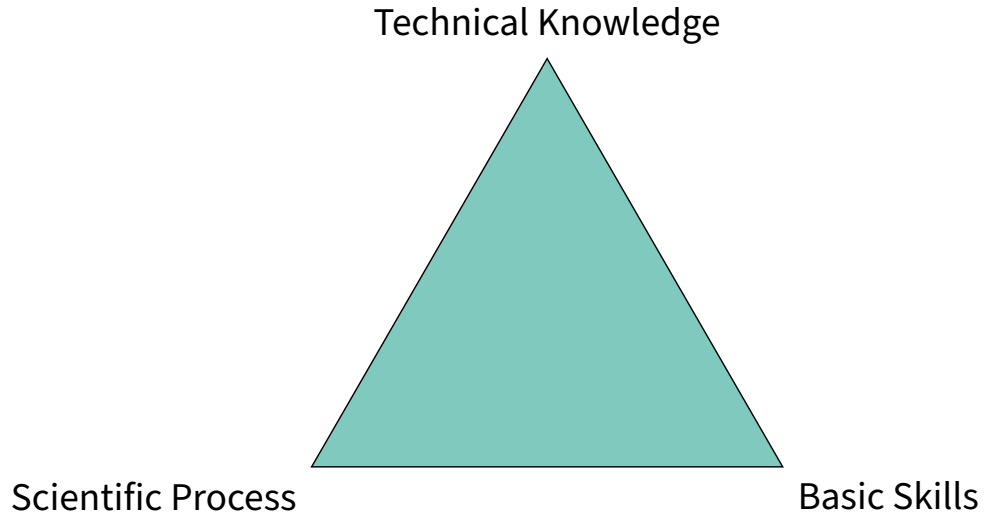


**Privacy und Technischer Datenschutz**  
**Seminar WS2023/24**  
**Basic Skills**

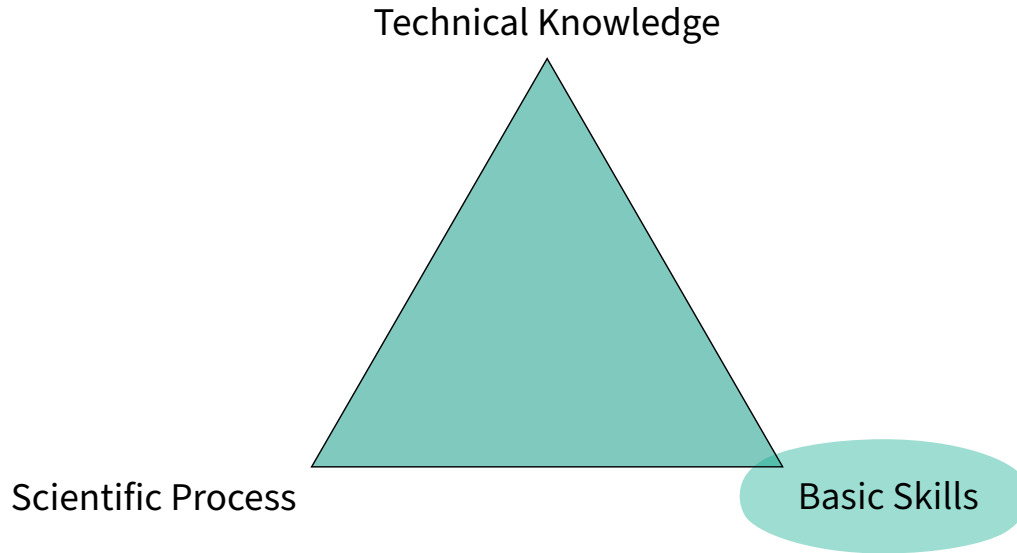
Patricia Guerra-Balboa

November 2, 2023

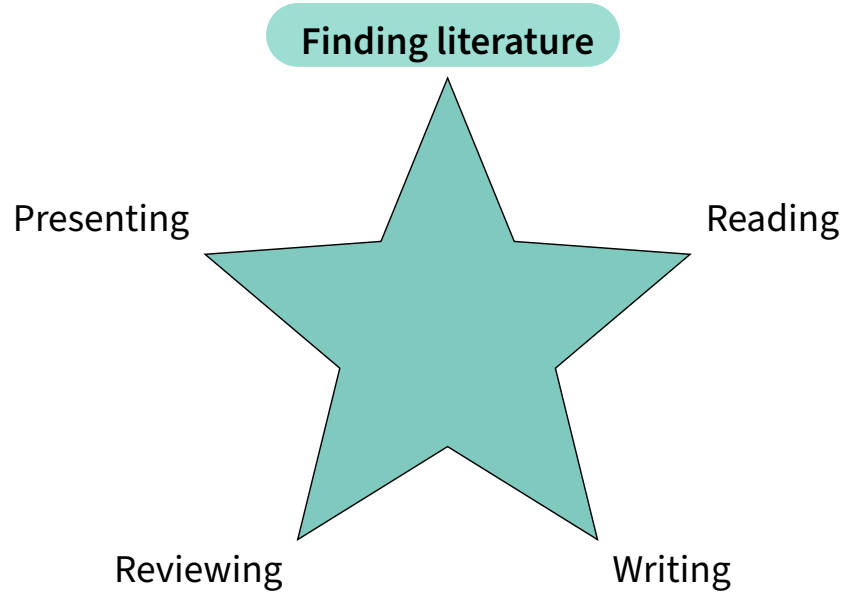
# Seminar goals



# Seminar goals



# Skills



# Finding literature

- ▶ Conferences/publication sites
- ▶ Search engines
  - ▶ Google Scholar
  - ▶ Springer
  - ▶ IEEE Xplore
  - ▶ DPBL
  - ▶ Citeseer
- ▶ ~~arXiv~~



# Search Techniques

## Backwards

Which papers are cited in the reference



## Forwards

Which papers cite the reference

**Figure 1:** The reference you are currently reading

# Search Techniques

## Backwards

Which papers are cited in the reference



## Forwards

Which papers cite the reference

**Figure 1:** The reference you are currently reading

# Finding literature






# Selection

## Check skim paper

- ▶ Area of research
- ▶ Assumptions, system vs. evaluation, . . .

- 
1. Title
  2. Abstract
  3. Conclusion
  4. Introduction
  5. Everything else (as needed)

## Check conference quality

- ▶ Ranking systems:
  - ▶ Core: A\*, A, B, C
  - ▶ (<http://portal.core.edu.au/conf-ranks/>)
  - ▶ ERA, Qualis,...
- ▶ Number of citations
- ▶ Year of publication



# Top Conferences

## ▶ (Practical) IT-Security:

A\* IEEE S&P (Security and Privacy)

Usenix NDSS (Network and Distributed System Security) Usenix Security

ACM CCS (Computer and Communications Security)

A : AsiaCCS, ESORICS, ...

## ▶ Privacy:

A PETS (Privacy Enhancing Technologies Symposium)

## ▶ Cryptography:

A\* Crypto (Advances in Cryptology ) EuroCrypt (Int. Conf. on the Theory and Application of Cryptographic Techniques)

A TCC, AsiaCrypt, FC,...

# Keep it organized

Reference management software

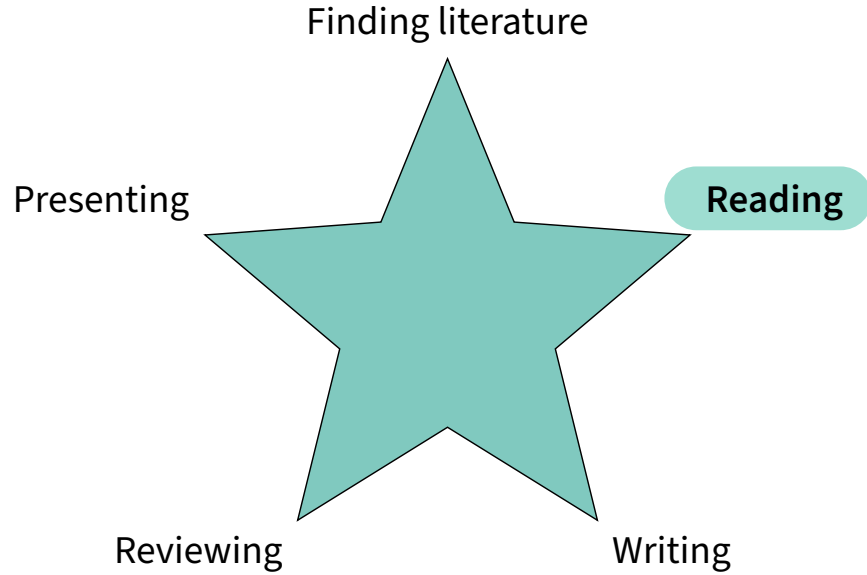
Zotero, Citavi, . . .

Tip:  
author+year+first\_word

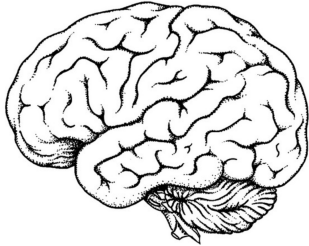


Example:  
Dwork2014algorithmic

# Skills



# Before Reading



Activate knowledge



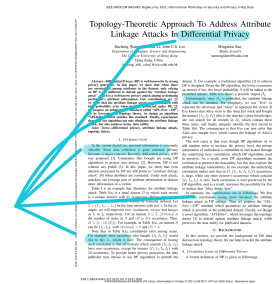
Guiding questions

# Techniques

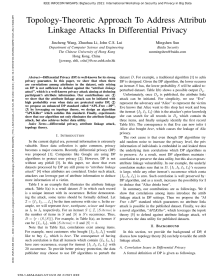
1. Title
2. Abstract
3. Conclusion
4. Introduction
5. Everything else (as needed)



skimming trough

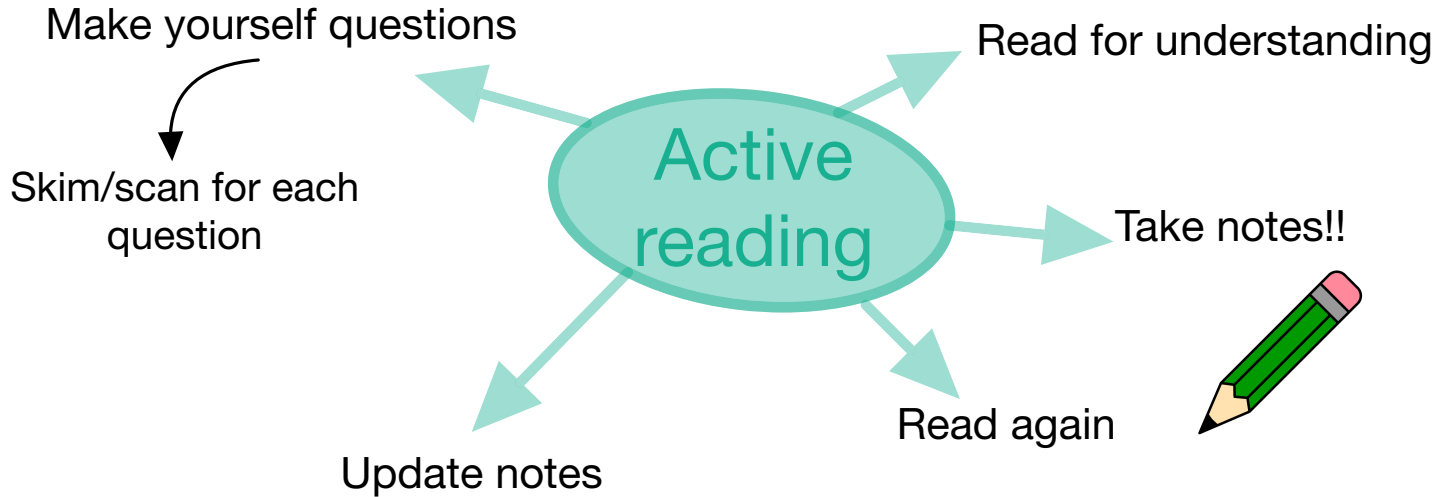


scanning



focused reading

# Possible reading strategy



# Further material on reading

- ▶ **“How to read a paper” by S. Keshav”**

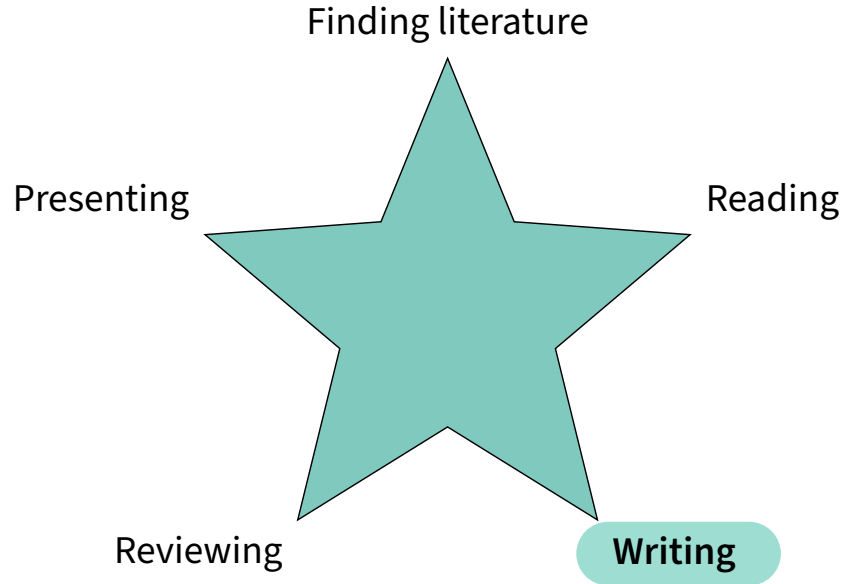
<http://blizzard.cs.uwaterloo.ca/keshav/home/Papers/data/07/paper-reading.pdf>

- ▶ **“About academic reading”**

<https://aso-resources.une.edu.au/academic-reading/about-academic-reading/>



# Skills



# Structure

- 0. Abstract
- 1. Introduction
- 2. Related work
- 3. Background
- 4. Main part
- 5. Conclusion & Future Work

IEEE INFOCOM WKSPs, BigSecurity 2021: International Workshop on Security and Privacy in Big Data

## Topology-Theoretic Approach to Address Attribute Linkage Attacks in Differential Privacy

Jincheng Wang, Zhenhua Li, John C.S. Lui  
Department of Computer Science and Engineering  
The Chinese University of Hong Kong  
Hong Kong, China  
{wangj16, lizh}@se.cuhk.edu.hk

Mingchen Sun  
Rutgers University  
sunmingch@hobbs.com

**Abstract**—Differential Privacy (DP) is well known for its strong privacy guarantee. In this paper, we show that when there are correlations among attributes in the dataset, only relying on DP is not sufficient to defend against the “attribute linkage attack”, which is a link-based privacy attack aiming at deducing a person’s sensitive information. Our contribution is: 1) we show that the attribute linkage attack can be initiated with high probability even when data are protected under DP; 2) we propose an enhanced DP standard called “APL-Free  $\epsilon$ -DP”, by leveraging our topology theory; we design an algorithm “APL-Killer” which satisfies the standard. Finally, experiments show that our algorithm not only eliminates the attribute linkage attack, but also achieves better data utility.

**I. INTRODUCTION**  
In the current digital era, personal information is extremely valuable. Since data collection is quite common, privacy becomes a major concern. Recently, differential privacy (DP) was proposed [1]. Companies like Google are using DP algorithms to protect user privacy [2]. However, DP is not without any pitfall [3]. In this paper, we show that most datasets processed by DP are still prone to “attribute linkage attack” [4] when attributes are correlated. Under such attack, attackers can leverage part of attribute information to deduce more information of a victim.

Table 1 is an example that illustrates the attribute linkage attack. Table 1(a) is a real dataset  $\mathcal{D}$  in which each record is a unique instant with its occurrence. Before demonstrating the attack, users should be formally defined. Let  $Z = \{z_1, \dots, z_N\}$  be the item universe with size  $N$ . In the example, we will represent user, activities, address and happen as  $z_1, z_2, z_3$  respectively. For an instant  $z \in Z$ ,  $\mathcal{X}(z)$  is the number of items in  $\mathcal{N}$  and  $\mathcal{S}(z)$  is  $\mathcal{N}$  occurrence. Thus  $\mathcal{Z} = \{z_1, \dots, z_N\}$ . For example, in Table 1(a), an instant  $z$  can be  $\{z_1, z_2\}$  with  $|\mathcal{X}(z)| = 2$  and  $|\mathcal{S}(z)| = 4$ .

Note that in Table 1(a), correlations exist among items. For example, most consumers who bought  $\{z_1, z_2, z_3\}$  would like to buy  $z_4$ , which is beer. The consequence of having such correlation is that all instants which contain  $\{z_1, z_2, z_3\}$  have zero occurrences for instant  $\{z_1, z_2, z_3, z_4\}$  with 30 occurrence. To provide better privacy protection, the data publisher may choose to use DP algorithms to perturb the

dataset  $\mathcal{D}$ . For example, a traditional algorithm [1] to achieve DP is to discard. Given the DP algorithm, the lower occurrence of instant  $z$  has the lower probability. It will be added to the perturbed dataset. Table 1(b) shows a possible output  $\mathcal{D}'$ . Unfortunately, even if  $\mathcal{D}'$  is published, the attribute linkage attack can be initiated. For simplicity, we use “Buy” to represent the arbitrary and “Alike” to represent the victim. If Eve knows that Alice went to this shop last week and bought the item  $\{z_1, z_2, z_3\}$ , then she knows Alice’s purchase history; she can search for all records in  $\mathcal{D}'$ , which contain these three items, and finally uniquely identify the first record in Table 1(b). The consequence is that Eve can now infer that Alice also bought beer, which causes the leakage of Alice’s privacy.

The root cause is that even though DP algorithms try to add random noise to attenuate the privacy level, the private information of individuals is embedded in and leaked through the underlying item correlations, such as DP algorithms need to preserve. As a result, most DP algorithms maintain the correlation to preserve the data utility, but this also exposes the attribute linkage vulnerability. In our example, the underlying correlation makes sure that in  $\mathcal{D}'$ ,  $\{z_1, z_2, z_3, z_4\}$  occurrence is large, while any other instant’s occurrence which contains  $\{z_1, z_2, z_3\}$  is zero. Such correlation is well preserved by the DP algorithm, and as a result, increases the probability for Eve to deduce that “Alice drinks beer”.

In summary, our contribution are as following. We first show that correlations among items introduce the attribute linkage attack in DP settings. Then we propose the “APL-Free  $\epsilon$ -DP” standard which guarantees no attribute linkage attack is possible in the published dataset. Finally, we design a novel algorithm, “APL-Killer”, which leverages the topology theory [5] to defend against attribute linkage attack, which improves the data utility of published datasets.

**II. BACKGROUND**  
In this section, we provide the background of DP, that includes how sampling theory [6] can help to reduce the attribute linkage attack.

A. Correlation Issues in Differential Privacy  
A formal definition of DP is given as follows.

IEEE INFOCOM WKSPs, BigSecurity 2021: International Workshop on Security and Privacy in Big Data

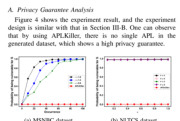


Fig. 4. Privacy guarantee analysis. (a) MSNDC dataset, (b) NCTCS dataset, (c) Check-Feedback dataset. The graphs plot relative error on the y-axis against the number of queries on the x-axis. APL-Killer (red line) maintains a significantly lower relative error compared to DP (blue line) and Pivots (green line) across all three datasets.

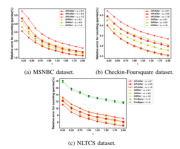


Fig. 5. Relative error for APL-Killer, DP, and Pivots. The graphs plot relative error on the y-axis against the number of queries on the x-axis. APL-Killer (red line) consistently shows the lowest relative error, followed by Pivots (green line), and DP (blue line) shows the highest relative error across all three datasets.

for MSNDC and Check datasets. In Figure 4 and Figure 5c, experiment results show that the relative error for APL-Killer is reduced by 1.0% in average, as a value. In Figure 5, one can check that APL-Killer reduces the relative error by 6.8% compared with that of DP, and 49% compared with that of Pivots. These show APL-Killer has a higher data utility. For traditional DP algorithms, although a smaller  $\epsilon$  can decrease the probability of being attacked, the data utility becomes worse. However, APL-Killer eliminates this dilemma. No matter how the privacy parameter  $\epsilon$  is set, the probability of being attacked is guaranteed to be zero. Therefore, our algorithm has the potential to publish the dataset with good data utility, while defending against the attribute linkage attack comprehensively.

**VI. CONCLUSIONS**  
In this paper, we show that the attribute linkage attack is a severe problem when using DP. In order to eliminate this attack, we improve DP and propose APL-Free  $\epsilon$ -DP. We further design an algorithm, APL-Killer, which leverages the topology-theoretic approach to defend against the attribute linkage attack. However, in our paper, we did not consider the probabilistic attribute linkage attack, which is a more advanced attack. Also, we did not set a clear restriction on how to choose APL-Killer’s parameters to get better data utility. These are potential directions for future research.

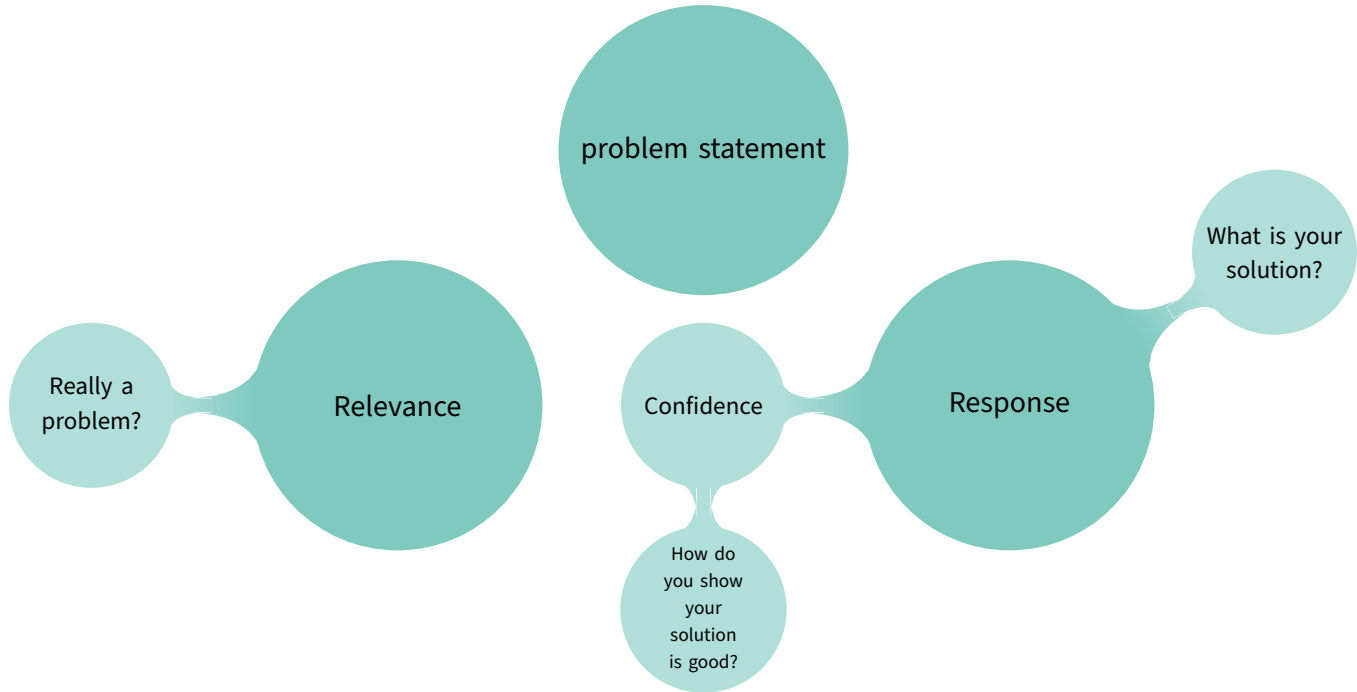
The work of John C.S. Lui was supported in part by the RGC GR032-18.

**REFERENCES**

- [1] D. Dworkin, et al., “The algorithmic foundations of differential privacy,” *Foundations and Trends in Theoretical Computer Science*, vol. 13, pp. 1–107, 2014.
- [2] J. News, “Differential privacy at scale: Uber and Berkeley collaboration,” in *Privacy 2017*, pp. 31–47, 2017.
- [3] M. Bun and S. Kulkarni, “On the impossibility of differentially private data release for unbounded domains,” *Proceedings of the ACM*, vol. 53, no. 4, pp. 2473–2484, 2020.
- [4] R. Chen, B. C. Pang, M. M. M. M., B. C. Pang, and K. Wang, “Privacy-preserving data publishing by local Laplace,” *Information Sciences*, vol. 253, pp. 61–71, 2013.
- [5] M. Sun, “Topology theory for network analysis,” *Journal of Mathematical Analysis and Applications*, vol. 443, no. 1, pp. 20–40, 2016.
- [6] M. Sun, “Topology of graph Laplacian matrices and its application in network analysis,” *Journal of Mathematical Analysis and Applications*, vol. 443, no. 1, pp. 20–40, 2016.
- [7] L. Sun, S. Chakrabarti, and P. Mittal, “Dependence index via correlation matrix for binary data,” *Journal of Machine Learning Research*, vol. 18, pp. 2018–2034, 2018.
- [8] R. Chen, S. M. M., B. C. Pang, B. C. Pang, and L. Sun, “Publishing unvalued data via differential privacy,” *Proceedings of the ACM*, vol. 53, no. 4, pp. 2473–2484, 2020.
- [9] M. Sun, “Topology theory for network analysis,” *Journal of Mathematical Analysis and Applications*, vol. 443, no. 1, pp. 20–40, 2016.
- [10] Technical report: <https://github.com/789638602/Topology-theoretic-approach>.



# Abstract

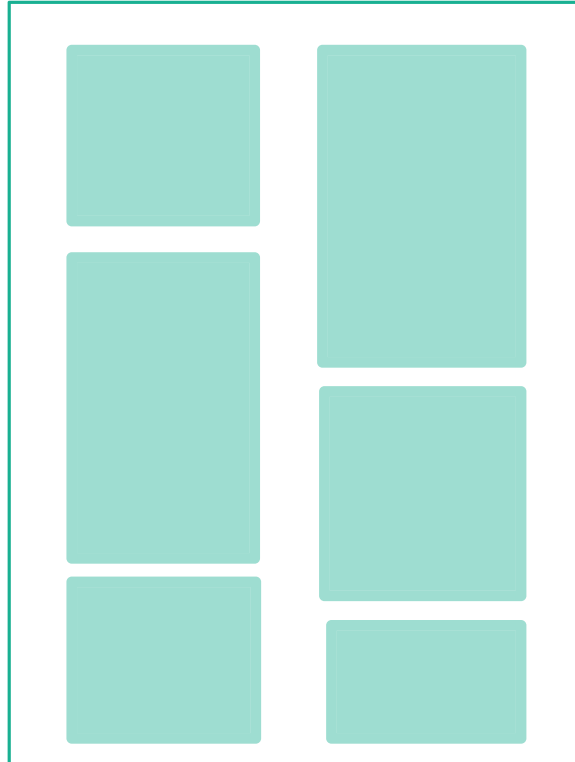


# Introduction

Broad topic  
& motivation

Specific topic  
&  
open problem

Goal  
&  
research question

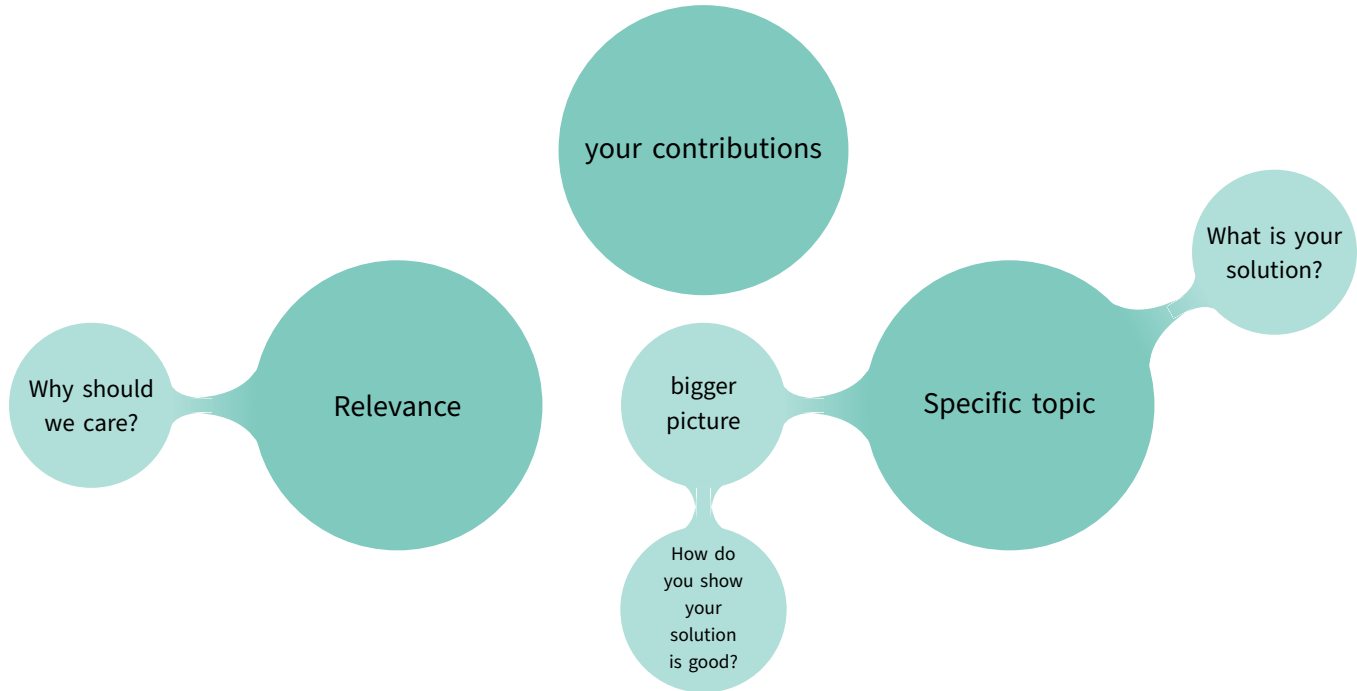


Scientific motivation  
&  
relevance

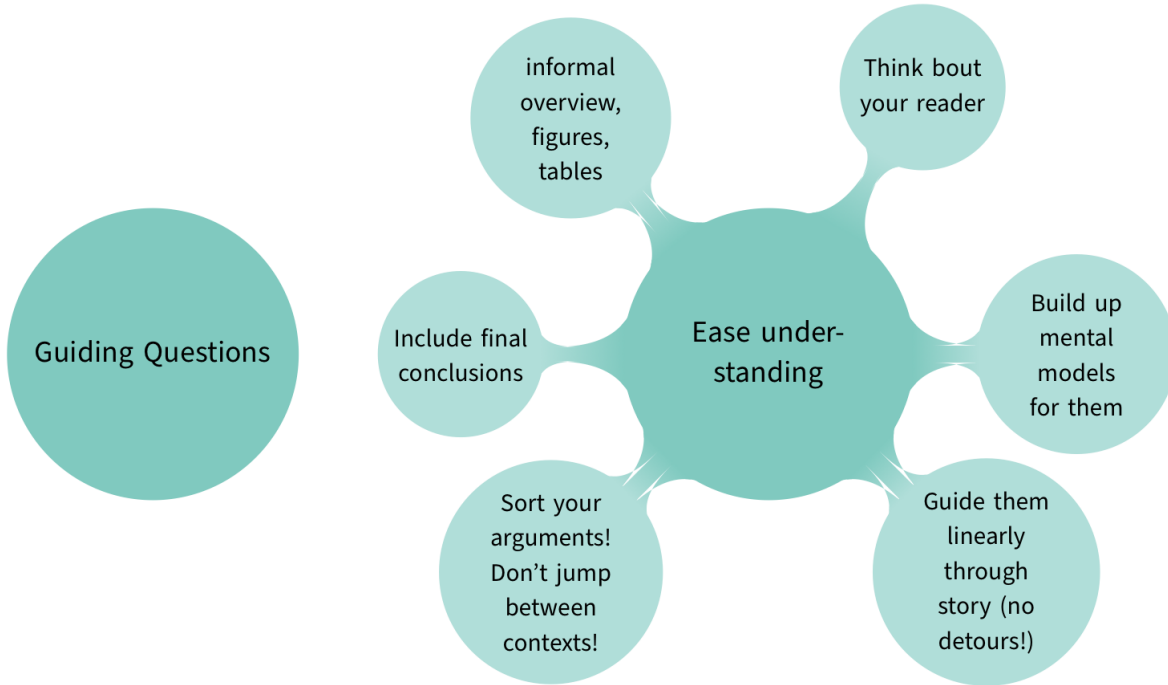
Your contributions

Reader's  
digest

# Conclusion



# main part



# Writing style

Basics: Grammar, spellcheck ...

## Scope:

- ▶ Sentence ↔ statement
- ▶ Paragraph ↔ idea
- ▶ Section ↔ subtopic

## KEEP IT SIMPLE!

- ▶ Short, precise sentences
- ▶ Active > passive
- ▶ Avoid negations
- ▶ Old → new

# Plagiarism

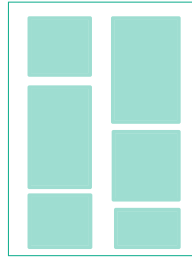
- ▶ Paraphrase: own words
  - ▶ Close your literature
- ▶ Signal:
  - ▶ Own content
  - ▶ Summary of someone else's
  - ▶ Direct quote



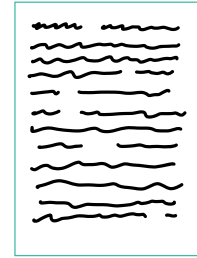
# How I approach it



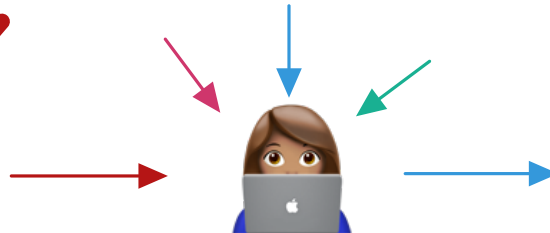
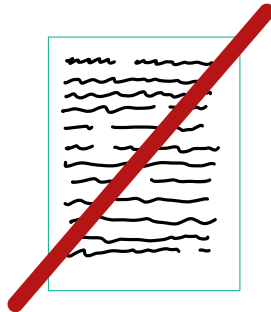
Rough plan



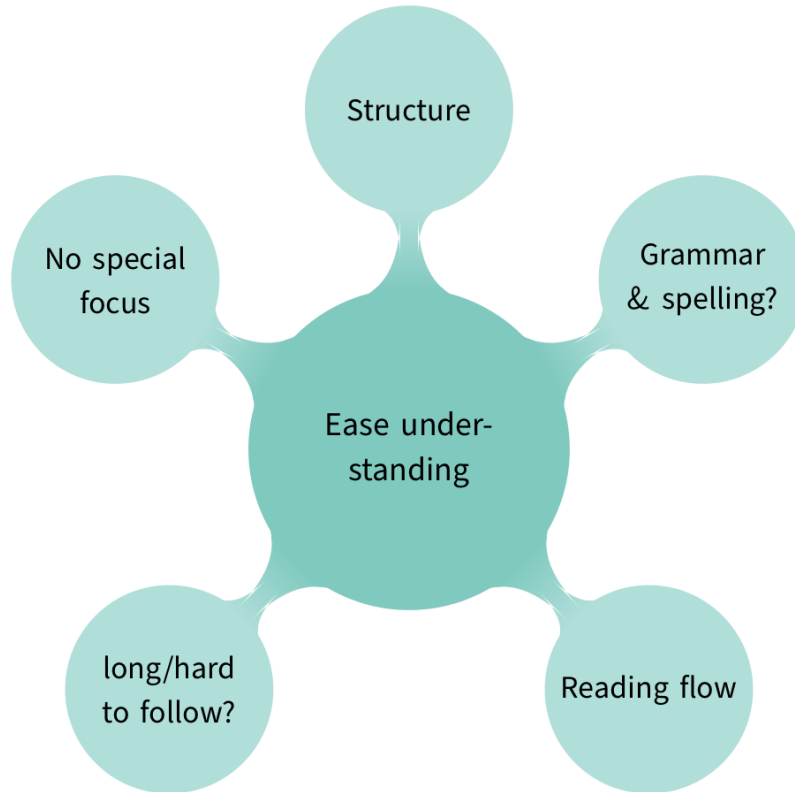
Structure



First draft



# Varying focus



# Further material on writing

- ▶ **“The Elements of Style” by Strunk and White**

<https://faculty.washington.edu/heagerty/Courses/b572/public/StrunkWhite.pdf>

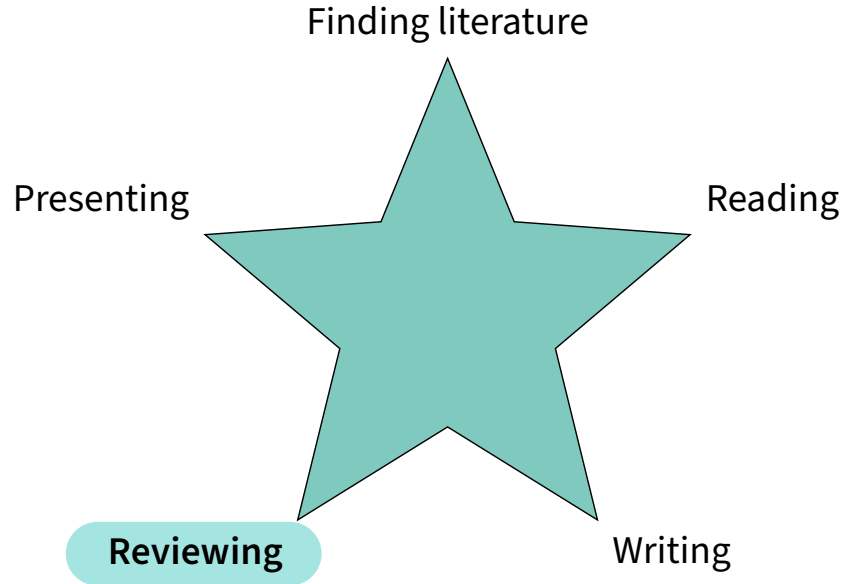
- ▶ **How to Write Papers So People Can Read Them:**

[https://www.youtube.com/watch?v=L\\_6xoMjFr70](https://www.youtube.com/watch?v=L_6xoMjFr70)

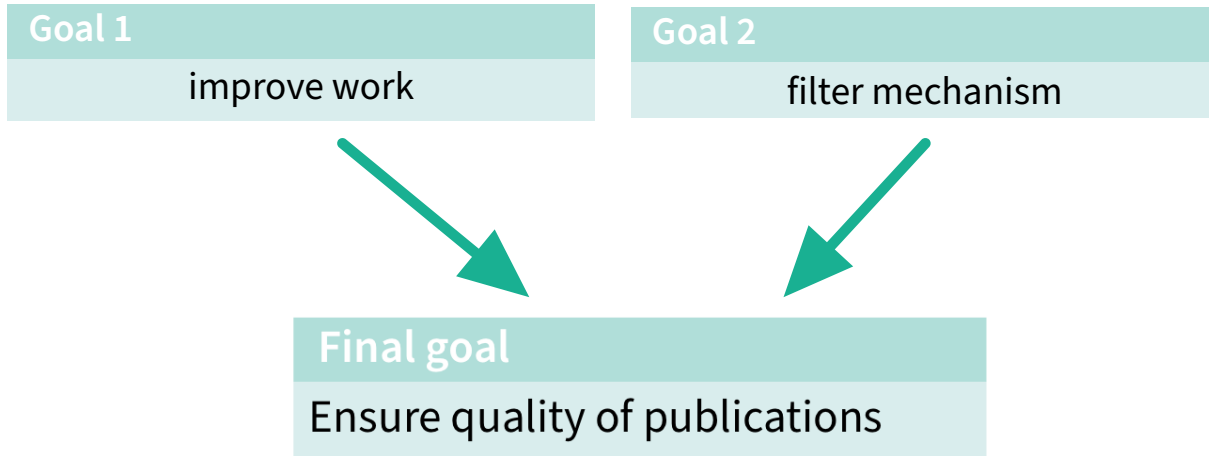
- ▶ **Plagiarism:**

[http://www.ou.edu/content/dam/integrity/docs/nine\\_things\\_you\\_should\\_know.pdf](http://www.ou.edu/content/dam/integrity/docs/nine_things_you_should_know.pdf)

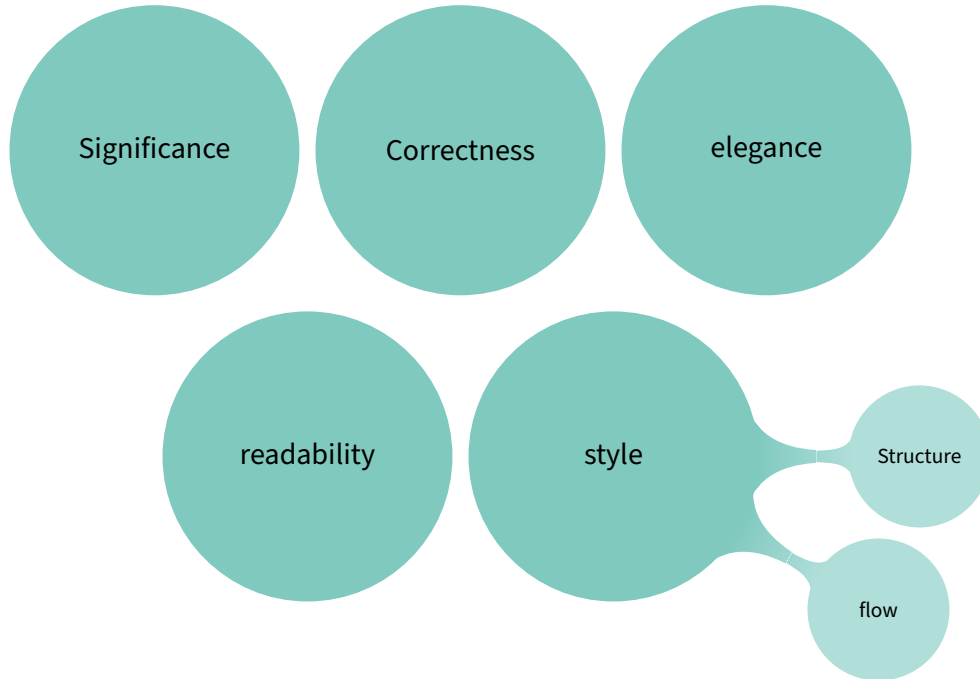
# Skills



# Why peer-reviewing?



# Quality criteria



# A good review

Thorough, critical

Follows given  
structure

Objective, polite

Helpful, con-  
structive, specific

anonymous

# Review Structure

- ▶ 3 Strengths & 3 Weaknesses
- ▶ Scale 1 – 5: each part of the paper:
  - ▶ Structure
  - ▶ Argumentation
  - ▶ Readability
  - ▶ Language
  - ▶ Grammar
  - ▶ Formatting
  - ▶ Citation Style
- ▶ Overall ranking (accept (strong/weak), reject(strong/weak))



# Opportunity: Receiving Reviews

Take your time for  
every point



Harsh/wrong/  
unfounded critics

Limited time

Learns what has been  
misunderstood

Open your mind

# Further material on Reviews

- ▶ **“The Task of the Referee”** by Alan Jay Smith:

<https://www.cs.utexas.edu/users/mckinley/notes/reviewing-smith.pdf>

- ▶ **“A Guide for New Referees in Theoretical Computer Science”** by Ian Parberry

[https://basics.sjtu.edu.cn/links/guide\\_referees.pdf](https://basics.sjtu.edu.cn/links/guide_referees.pdf)

# Skills



# Purpose first!

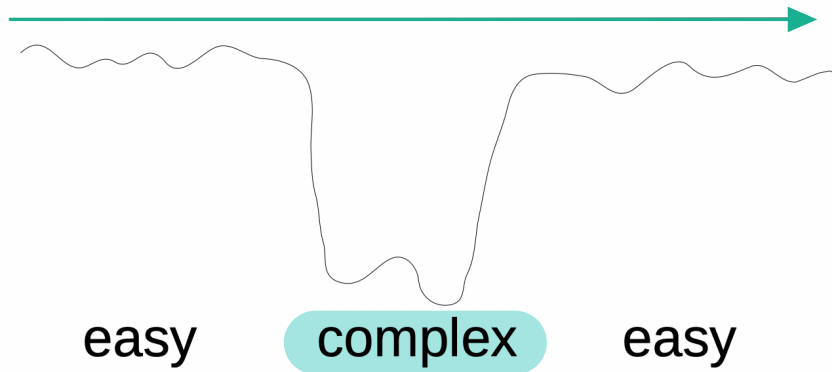


PERSUADE

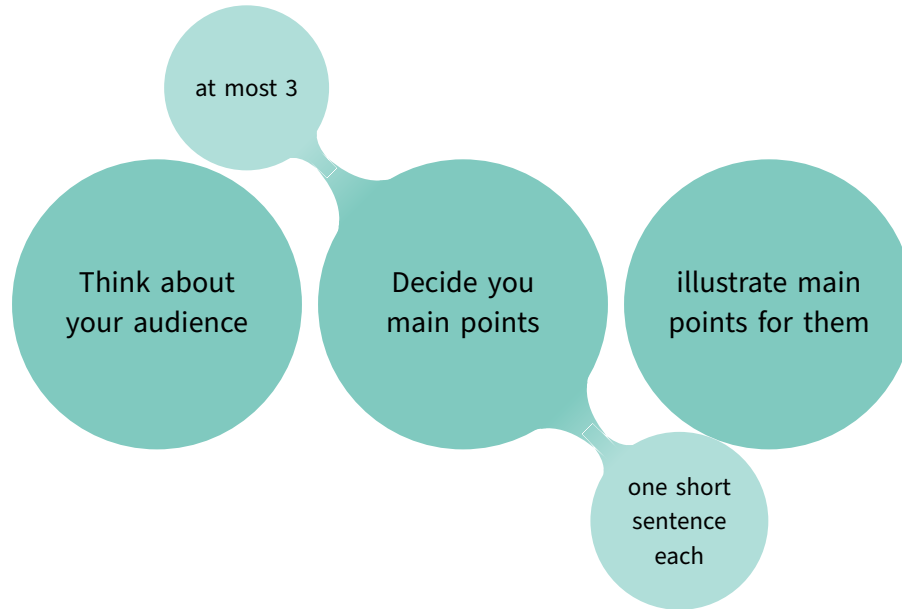
INFORM

ENTERTAIN

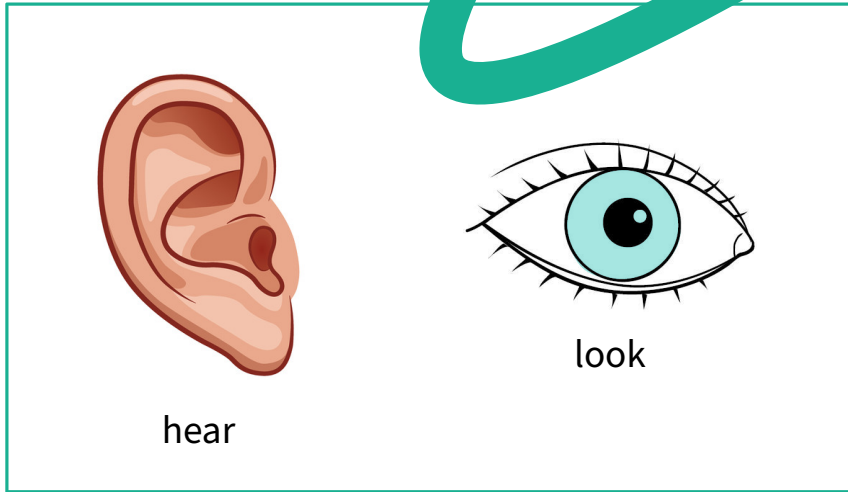
# The grebe strategy



# building the presentation strategy



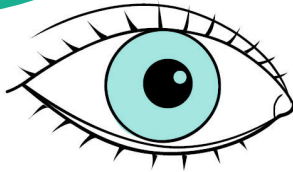
# The basics



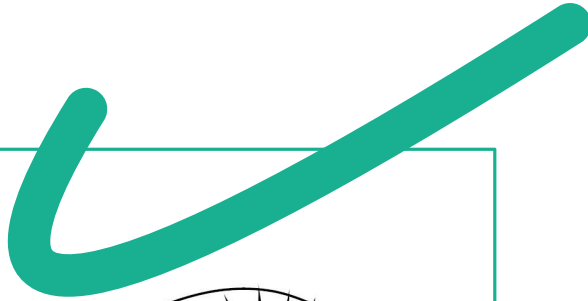
# The basics



hear



look




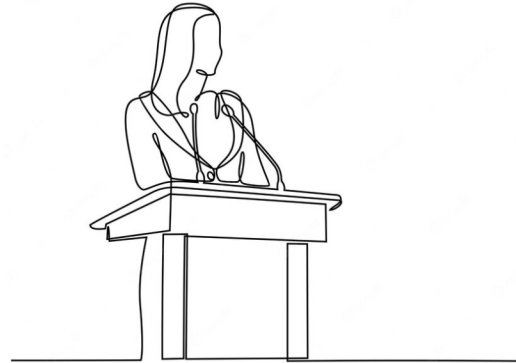
DO NOT READ

Figures ↑↑ Vs. Text ↓↓




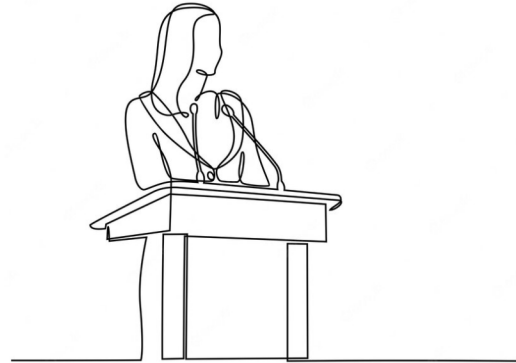
# The basics

- ▶ Do not read! 
- ▶ Look to the people
- ▶ Use your body language
- ▶ Change your voice



# The basics

- ▶ Do not read! 
- ▶ Look to the people
- ▶ Use your body language
- ▶ Change your voice



Slow

Fast

# Not To Do List

- ▶ Not signaling own/other's contributions
- ▶ Finish after 2/3 of the allowed time
- ▶ Go 1/3 over time
- ▶ Include everything - all the details!
- ▶ Cover every part, but give no details at all (No depth)
- ▶ Only cover a tiny part of your work (No breadth)

# Further material on presenting

- ▶ **“How to avoid death By PowerPoint”** by David JP Phillips:  
<https://www.youtube.com/watch?v=Iwpi1Lm6dFo>
- ▶ **“PowerSpeak”** by Dorothy Leeds

Good luck!

