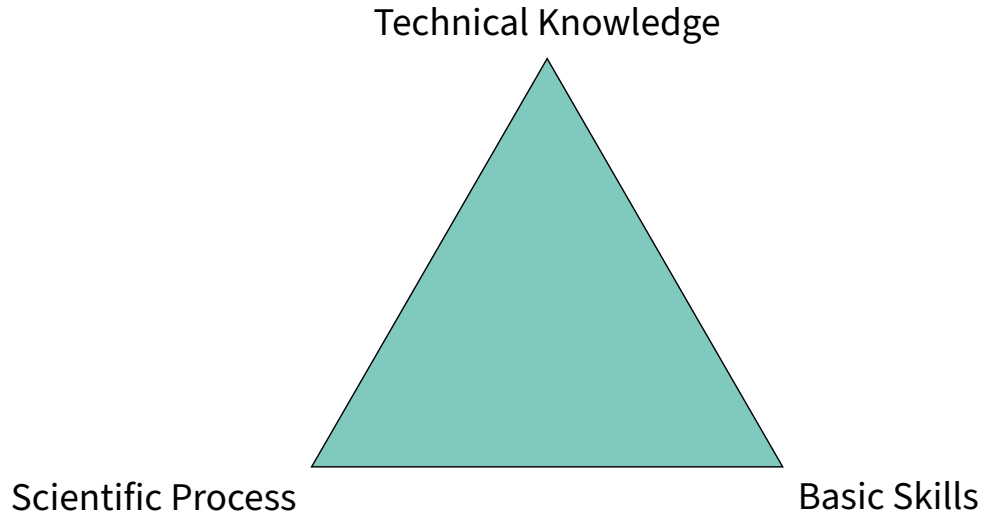


Privacy und Technischer Datenschutz
Seminar SS2023
Basic Skills

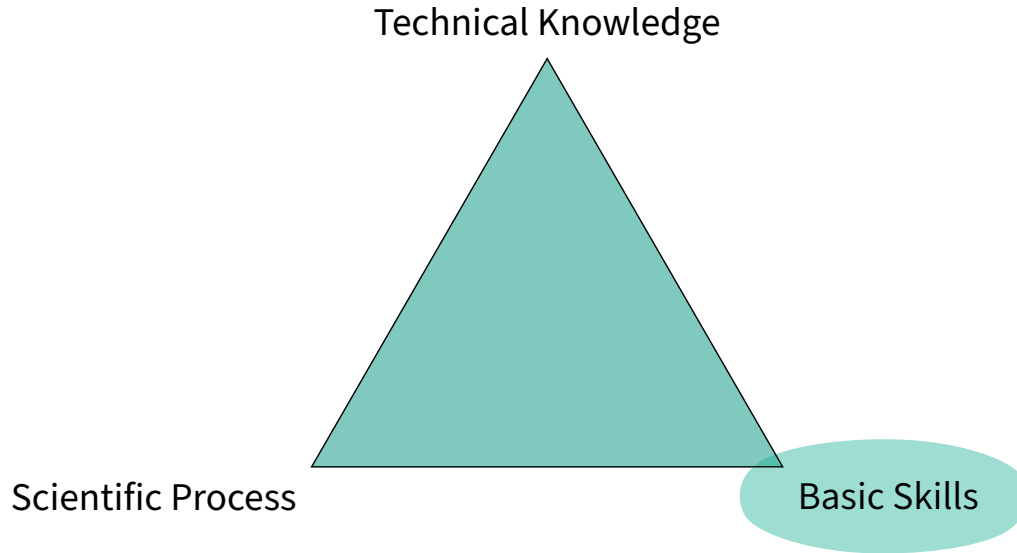
Patricia Guerra-Balboa

April 25, 2023

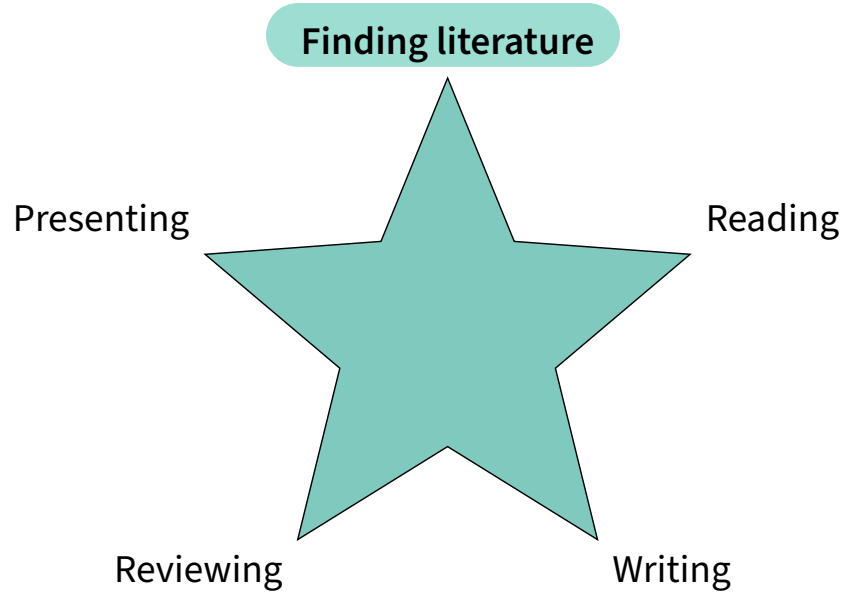
Seminar goals



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

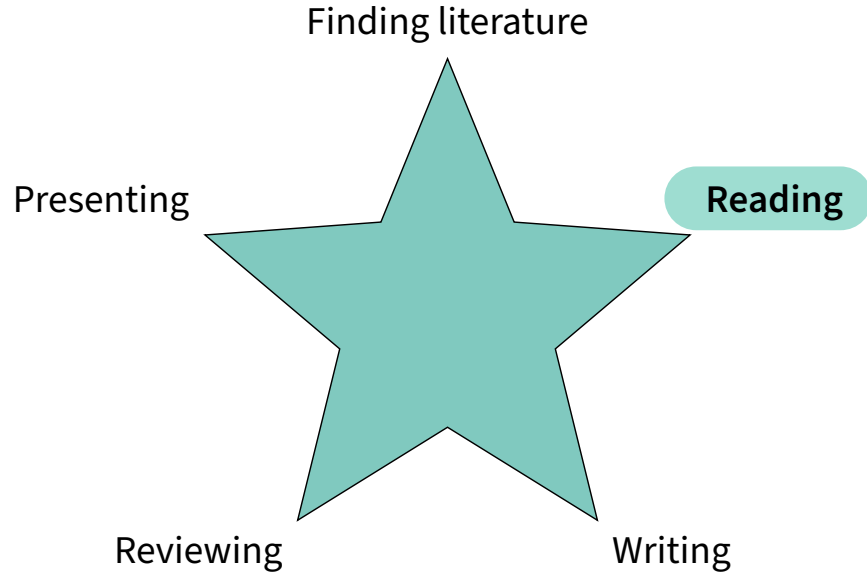
Zotero, Citavi, . . .

Tip:
author+year+first_word

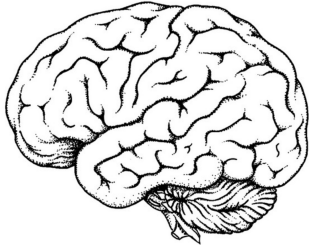


Example:
Dwork2014algorithmic

Skills



Before Reading



Activate knowledge



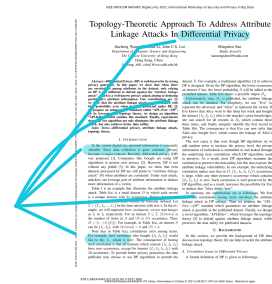
Guiding questions

Techniques

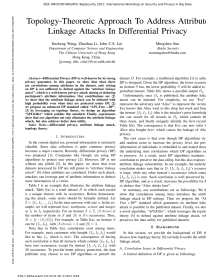
1. Title
2. Abstract
3. Conclusion
4. Introduction
5. Everything else (as needed)



skimming trough

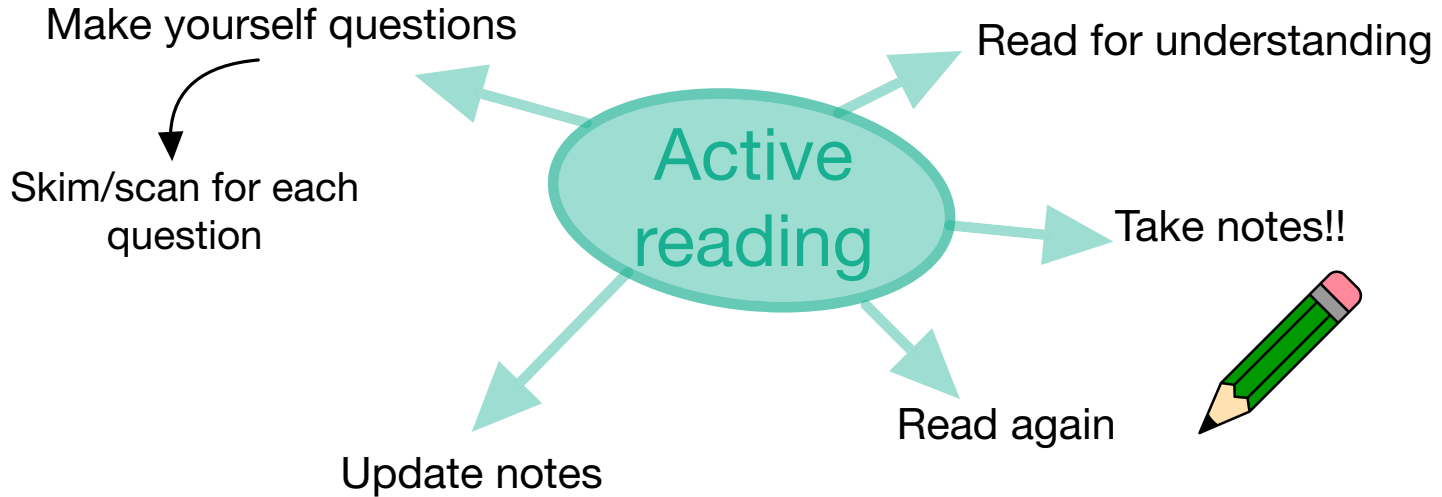


scanning



focused reading

Possible reading strategy



Further material on reading

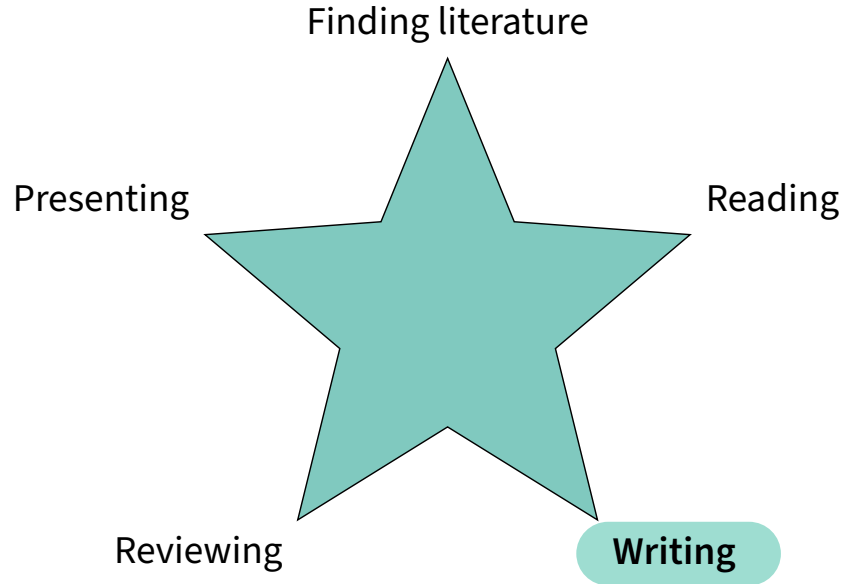
- ▶ **“How to read a paper” by S. Keshav”**

<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, Zhonghua 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
sunmingc@rutgers.edu

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 ϵ -DP”, by leveraging our topology theory; we design an algorithm “APL-Killer” which satisfies the standard. Finally, experiments show that our algorithm can only eliminate 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 D in which each record is a unique instant with its occurrence. Before disseminating 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 users, activities, activities and happen as z_1, z_2 respectively. For an instant $z_i \in Z$, $\mathcal{I}(z_i)$ is the number of items in N and $\mathcal{I}(z_i) \geq 3$ occurrence. Thus $\mathcal{I}(z_1) = \{1, 2, 3, 4, 5\}$. For example, in Table 1(a), an instant z_1 can be $\{1, 2, 3, 4, 5\}$ with $|\mathcal{I}(z_1)| = 5$ and $|\mathcal{I}(z_2)| = 4$.

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

dataset D . For example, a traditional algorithm [1] to achieve DP is to discard. Given the DP algorithm, the lower occurrence of instant z_4 has the lower probability. It will be added to the perturbed dataset. Table 1(b) shows a possible output D' . Unfortunately, even if D' is published, the attribute linkage attack can be initiated. For simplicity, we use “Buy” to represent the adversary and “Alice” 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 can infer Alice’s purchase history. She can search for all records in 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 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 instance correlation as in following, we first show that correlations among items introduce the attribute linkage attack in DP settings. Then we propose the “APL-Free ϵ -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, then we describe how topology theory [5] can help to tackle 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

A. Privacy Guarantee Analysis

Figure 4 shows the experiment result, and the experiment design is similar with that in Section III-B. One can observe that by using APL-Killer, there is no single APL in the generated dataset, which shows a high privacy guarantee.



Fig. 4. Privacy comparison using real-world datasets.

B. Data Utility Analysis

Since counting query is the most fundamental operation in data mining today, we focus on it. For each parameter setting, 50,000 random counting queries are generated. Given a query Q , the relative error $RE(Q)$ is computed as $\frac{|Q(D') - Q(D)|}{|Q(D)|}$, where $Q(D')$ is the query result on the generated dataset D' and $Q(D)$ is the query result on the original dataset, and it is the ratio between the number of records in D' and D . The smaller the relative error, the better the data utility. Since we are equally bound in order to weaken the influence of queries with extremely small counting answers, we set the ratio bound to 100% of the size of the original dataset.

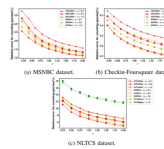


Fig. 5. Relative error for APL-Killer, DP+P and PivRays.

In Figure 5, each point in the average computed by generating 50,000 queries in terms of 1,000 records. For DP+P and APL-Killer, we change the parameter ϵ and c_0 , which is used to control the partitioning bound. In APL-Killer, we allow users to set C for generating histograms, and we set C' to c_0 , that any length l . Note that it is time-consuming for PivRays to process large datasets (over 24 hours), so PivRays is not used

for MSNBC and Checkin datasets. In Figure 5(a) and Figure 5(b), experiment results show that the relative error for APL-Killer is reduced by 1.0% in average, as a value. In Figure 5(c), one can check that APL-Killer reduces the relative error by 6.8% compared with that of DP+P, and 49% compared with that of PivRays. These show APL-Killer has a higher data utility. For traditional DP algorithms, although a smaller ϵ can decrease the probability of being attacked, the data utility becomes worse. However, APL-Killer eliminates this dilemma. No matter how the privacy parameter ϵ is set, the probability of being attacked is guaranteed to be zero. Therefore, our algorithm lets publishers 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 ϵ -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. There are potential directions for future research.

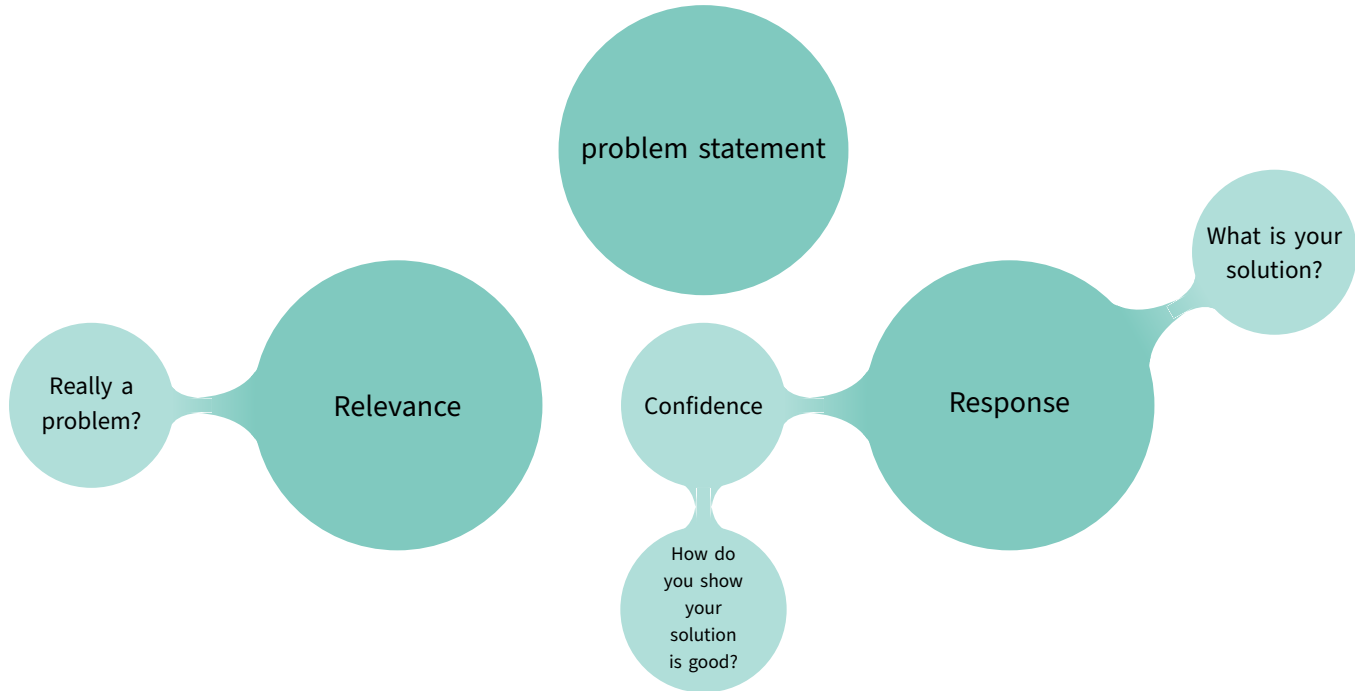
The work of John C.S. Lui was supported in part by the RGC GR012-18.

REFERENCES

- [1] D. Dworkin, “A firm foundation for differential privacy,” *Proceedings and Foundations of Theoretical Computer Science*, pp. 409–425, 2006.
- [2] J. News, “Differential privacy at scale: Uber and Berkeley collaboration,” in *Privacy 2018: Designing Privacy for the Real World*, pp. 1–12, 2018.
- [3] M. Bun, “Differential privacy: A practical guide to modern data protection,” in *International Conference on Database Theory*, pp. 1–23, 2017.
- [4] B. Chen, B. C. Pang, M. McMichael, B. C. Chen, and K. Wang, “Differential privacy: A practical guide to modern data protection,” in *International Conference on Database Theory*, pp. 1–23, 2017.
- [5] C. Li, “Topology theory for network analysis,” *Journal of mathematical analysis and applications*, vol. 331, pp. 481–501, 2007.
- [6] M. Feldman, “The art of data: Linking privacy, utility, and accuracy,” in *Proceedings of the ACM*, vol. 2017, pp. 1–12, 2017.
- [7] S. Dworkin, “A firm foundation for differential privacy,” in *Proceedings of the ACM*, vol. 2006, pp. 409–425, 2006.
- [8] B. Chen, M. McMichael, B. C. Pang, B. C. Chen, and L. Xing, “Differential privacy at scale: Uber and Berkeley collaboration,” in *Privacy 2018: Designing Privacy for the Real World*, pp. 1–12, 2018.
- [9] B. Chen, B. C. Pang, M. McMichael, B. C. Chen, and K. Wang, “Differential privacy: A practical guide to modern data protection,” in *International Conference on Database Theory*, pp. 1–23, 2017.
- [10] Technical report <https://github.com/berkeleylab/2018-02-02-2017>.



Abstract

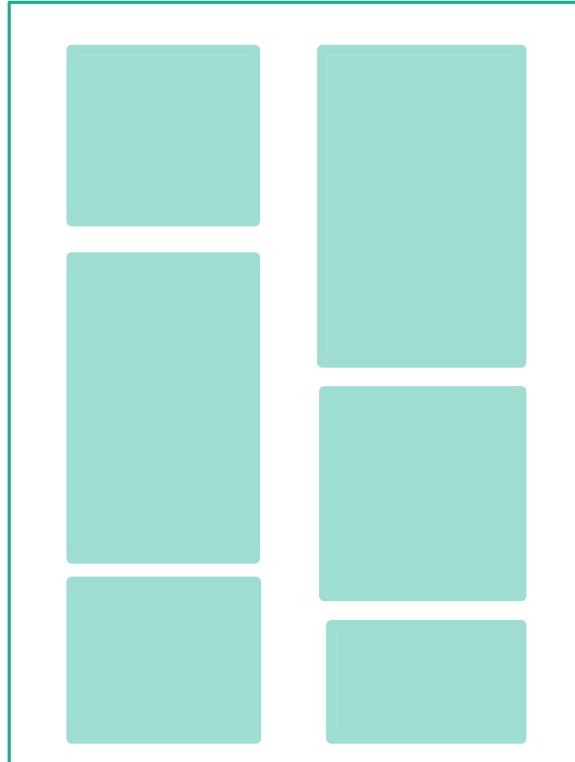


Introduction

Broad topic
& motivation

Specific topic
&
open problem

Goal
&
research question

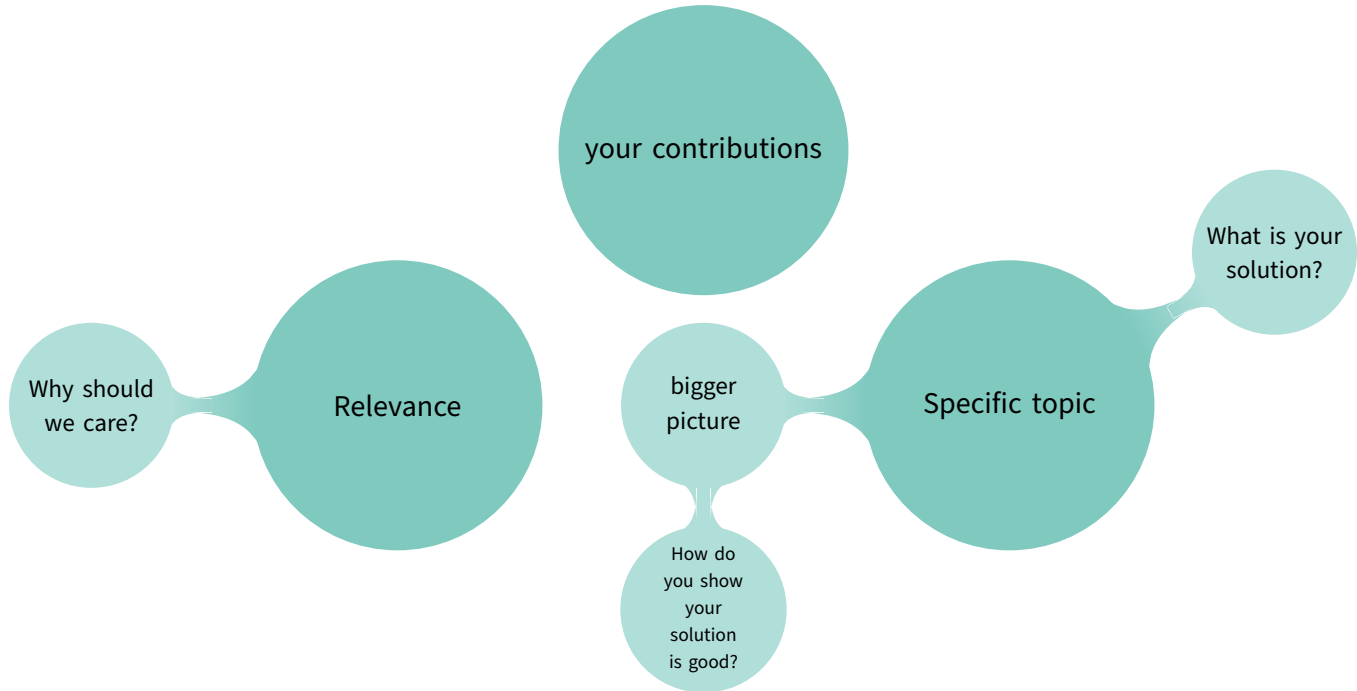


Scientific motivation
&
relevance

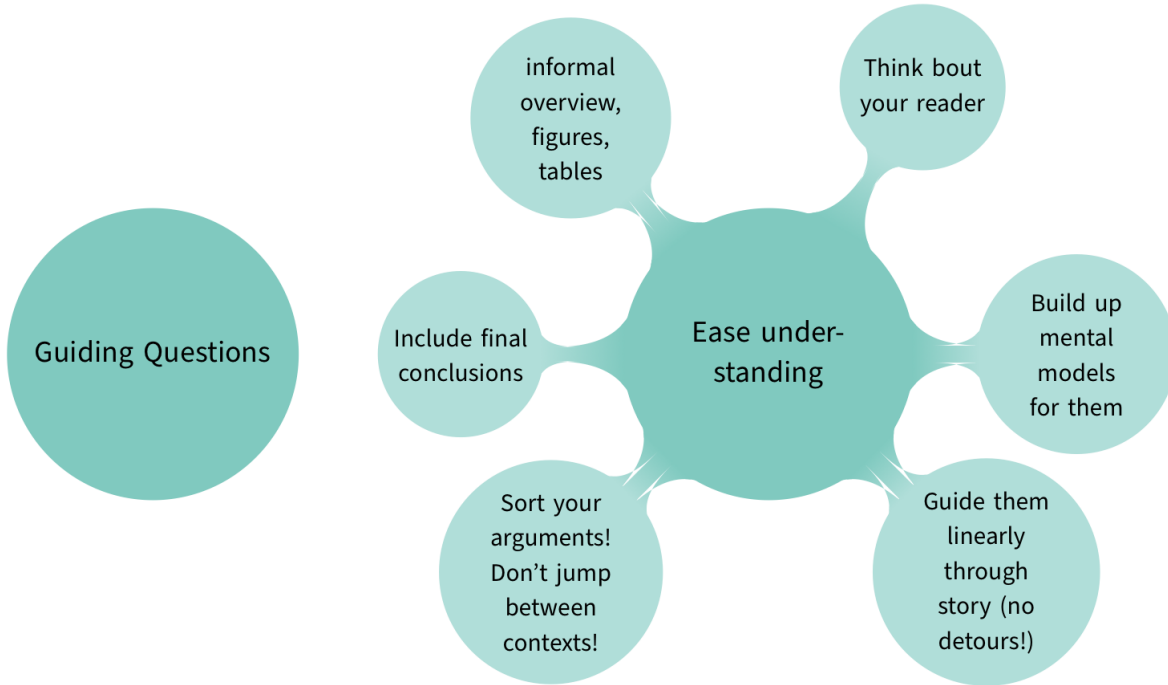
Your contributions

Reader's
digest

Conclusion



main part



Writing style

Basics: Grammar, spellcheck ...

Scope:

- ▶ Sentence \leftrightarrow statement
- ▶ Paragraph \leftrightarrow idea
- ▶ Section \leftrightarrow subtopic

KEEP IT SIMPLE!

- ▶ Short, precise sentences
- ▶ Active $>$ passive
- ▶ Avoid negations
- ▶ Old \rightarrow new

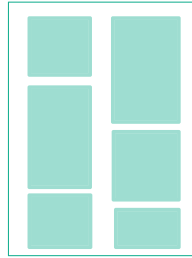
Plagiarism

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

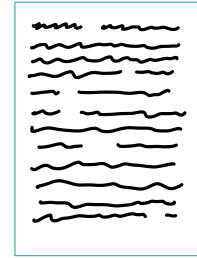
How I approach it



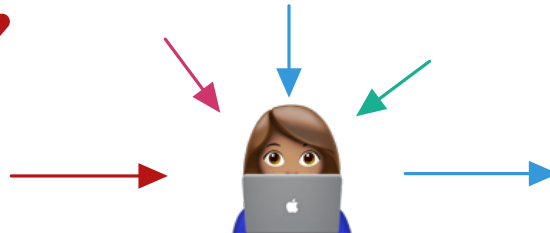
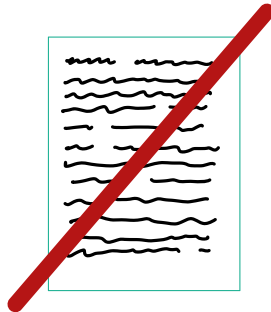
Rough plan



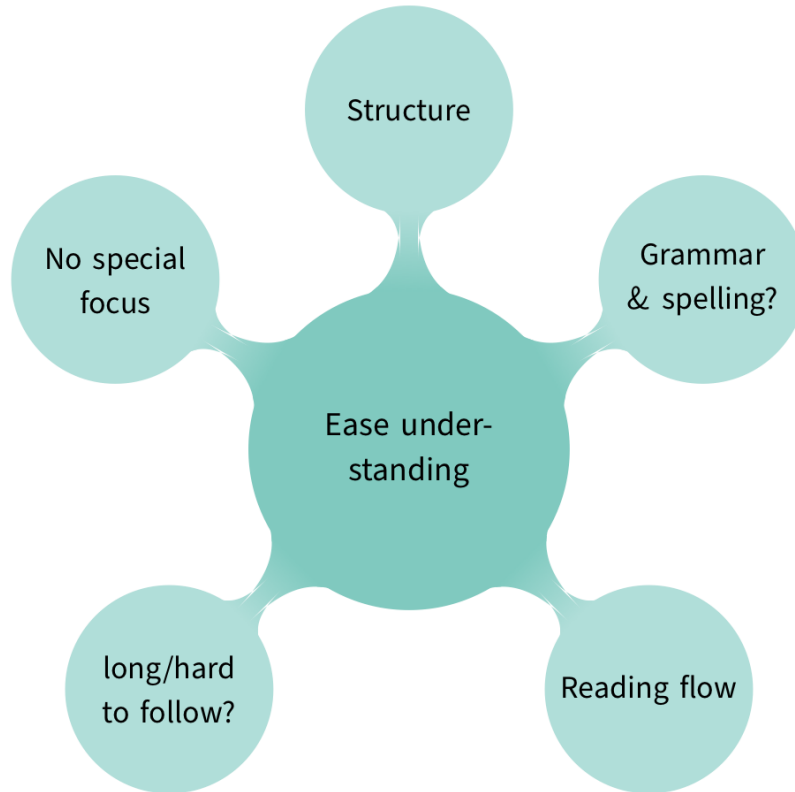
Structure



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>

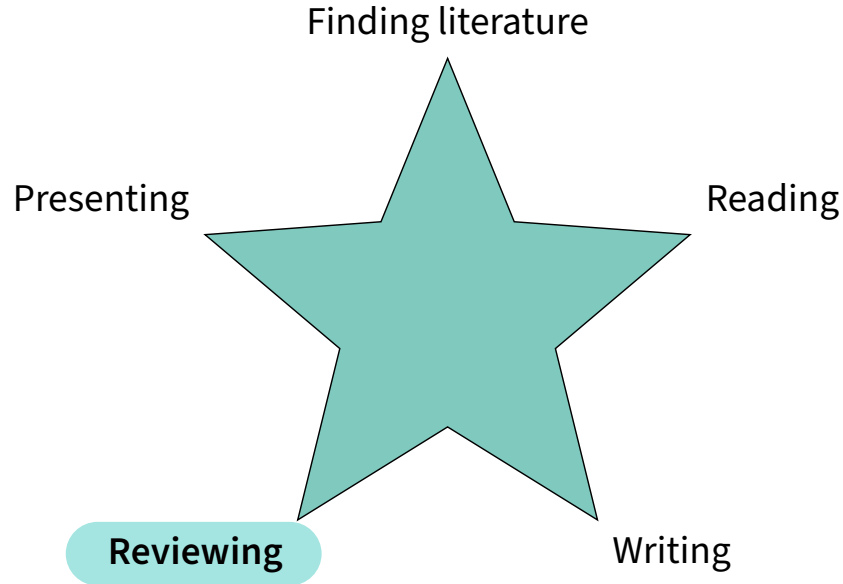
- ▶ **How to Write Papers So People Can Read Them:**

https://www.youtube.com/watch?v=L_6xoMjFr70

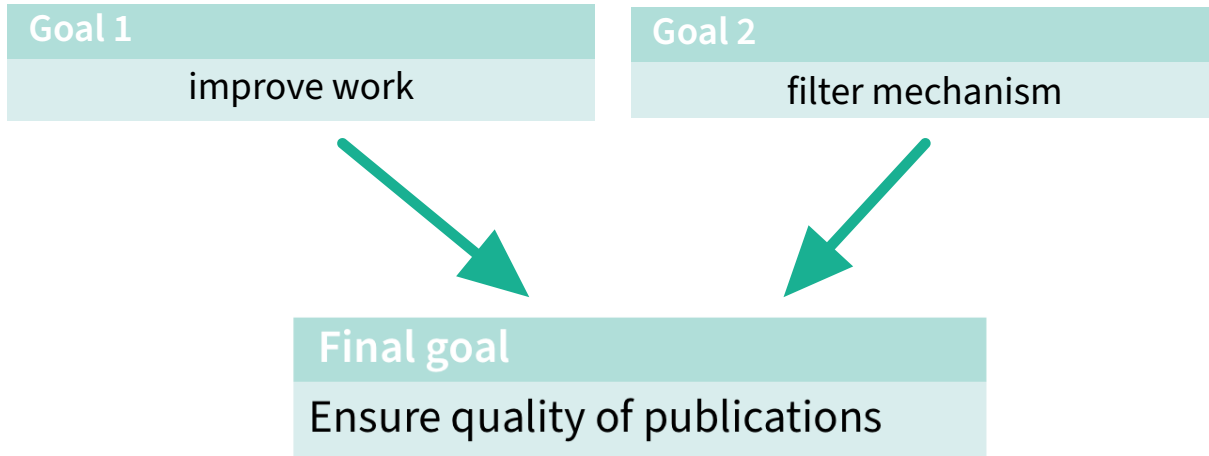
- ▶ **Plagiarism:**

http://www.ou.edu/content/dam/integrity/docs/nine_things_you_should_know.pdf

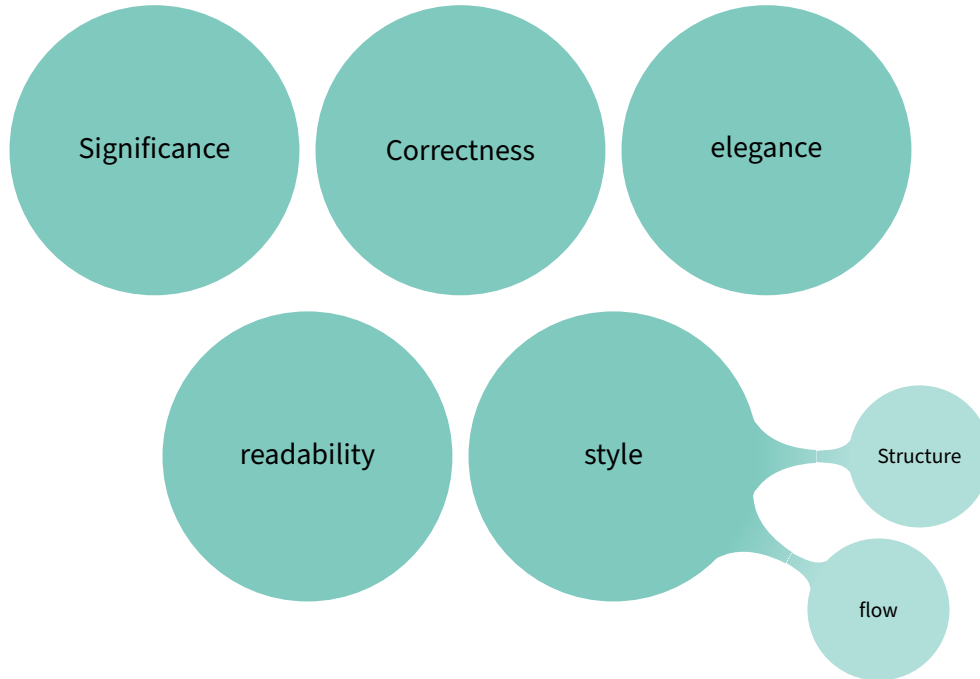
Skills



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

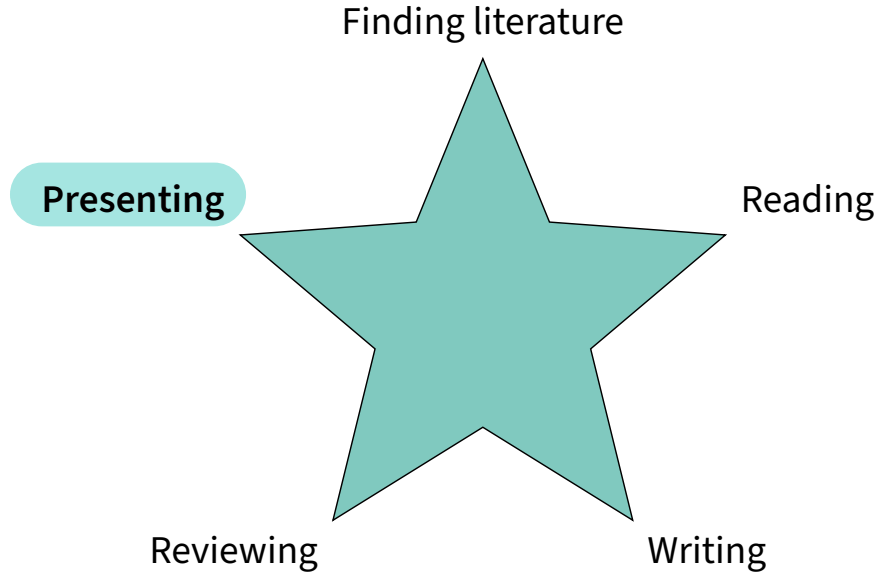
Learns what has been
misunderstood

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

Skills



Purpose first!

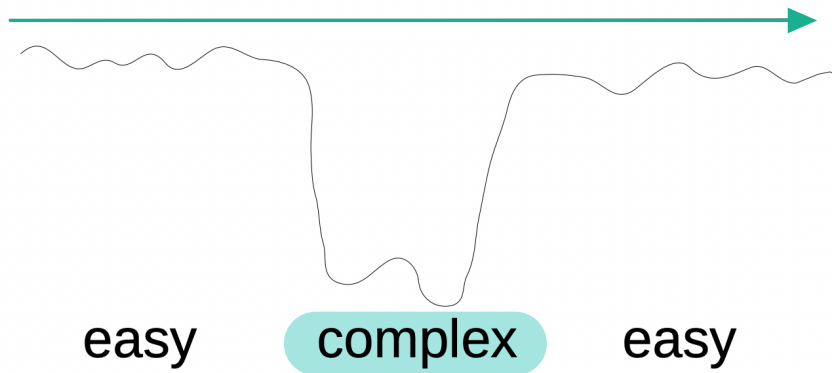


PERSUADE

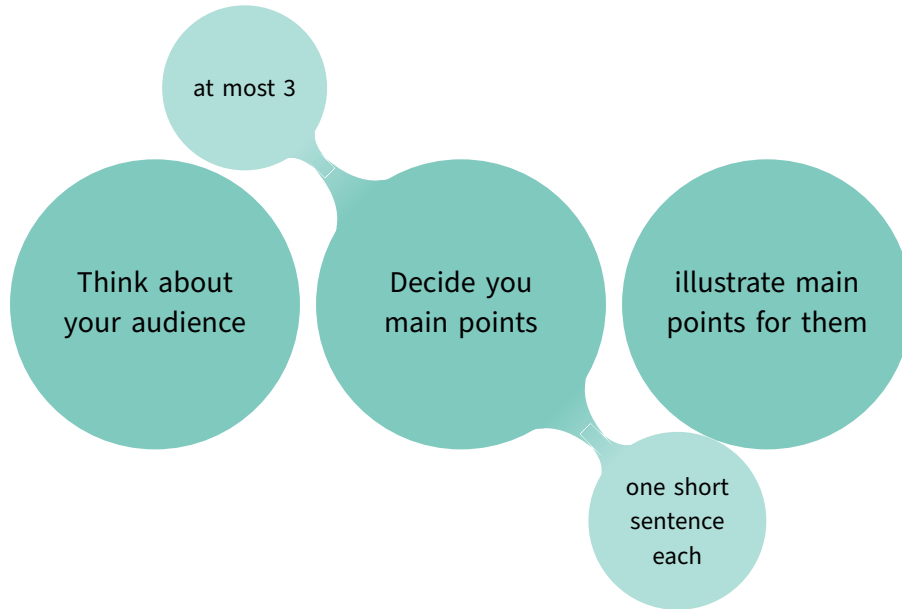
INFORM

ENTERTAIN

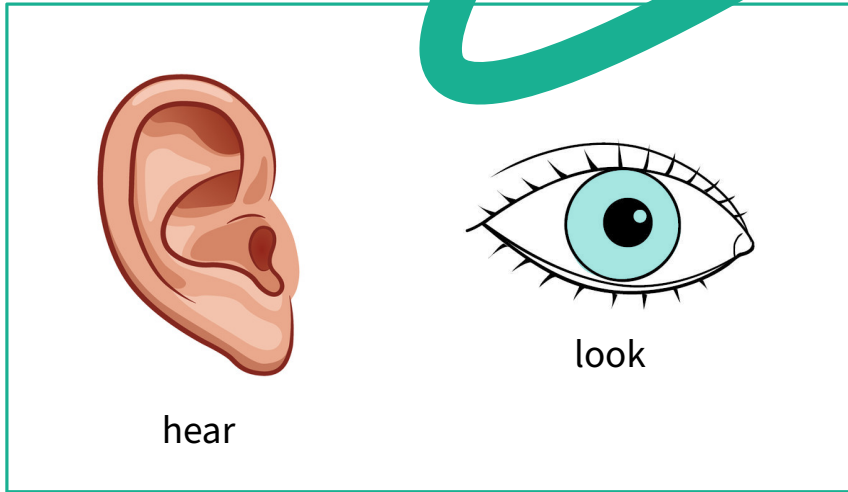
The grebe strategy



building the presentation strategy



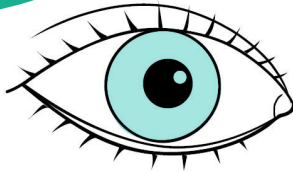
The basics



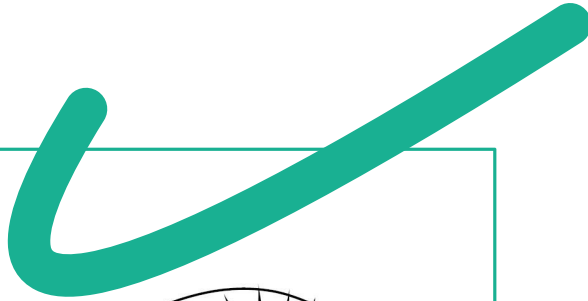
The basics



hear




look

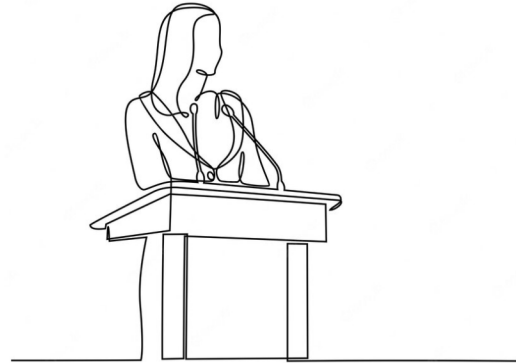


DO NOT READ


Figures ↑↑ Vs. Text ↓↓

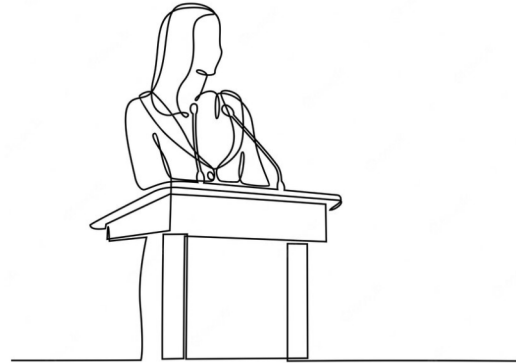
The basics

- ▶ 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!

