IBM HR EMPLOYEE ATTRITION MANAGEMENT

ATTRITION PREDICTION

3X DATA MINING GROUP MEMBERS

IGOR PEDEVANI, MOHAMMED TOPIWALLA ,PATRICIA LONDONO, MILLICENT OTCHERE, SANCHITA KUMARI, VALERIO TROTTA

Problem Statement

- Our client is IBM a leading firm and in the IT sector. It is recently facing a steep increase in its employee attrition . Employee attrition has gone up from 14% to 25% in the last 1 year . We are asked to prepare a strategy to immediately tackle this issue such that the firm's business is not hampered and also to propose an efficient employee satisfaction program for the long run. Currently, no such program is in place . Further salary hikes are not an option.
- The attrition problem is not only unique to ibm but to other IT companies such as Infosys, India's second largest IT services company, that is also battling high attrition, with a peak attrition of 20.4 % in the October-December quarter of FY15.

HOW CAN WE REDUCE IBM COMPANY'S ATTRITION RATE BY PREDICTING IF A CANDIDATE WILL EXIT IN INDIA WITHIN THE YEAR?

- **S**pecific :- To Indian geography in IBM
- **M**easurable:- To reduce attrition rate(By at least 5%)
- **A**ction oriented:- Reduce employee attrition & suggest employee engagement & satisfaction programs
- **R**elevant:- Direct impact on company's top and bottom line
- **T**ime bound :- 12 months

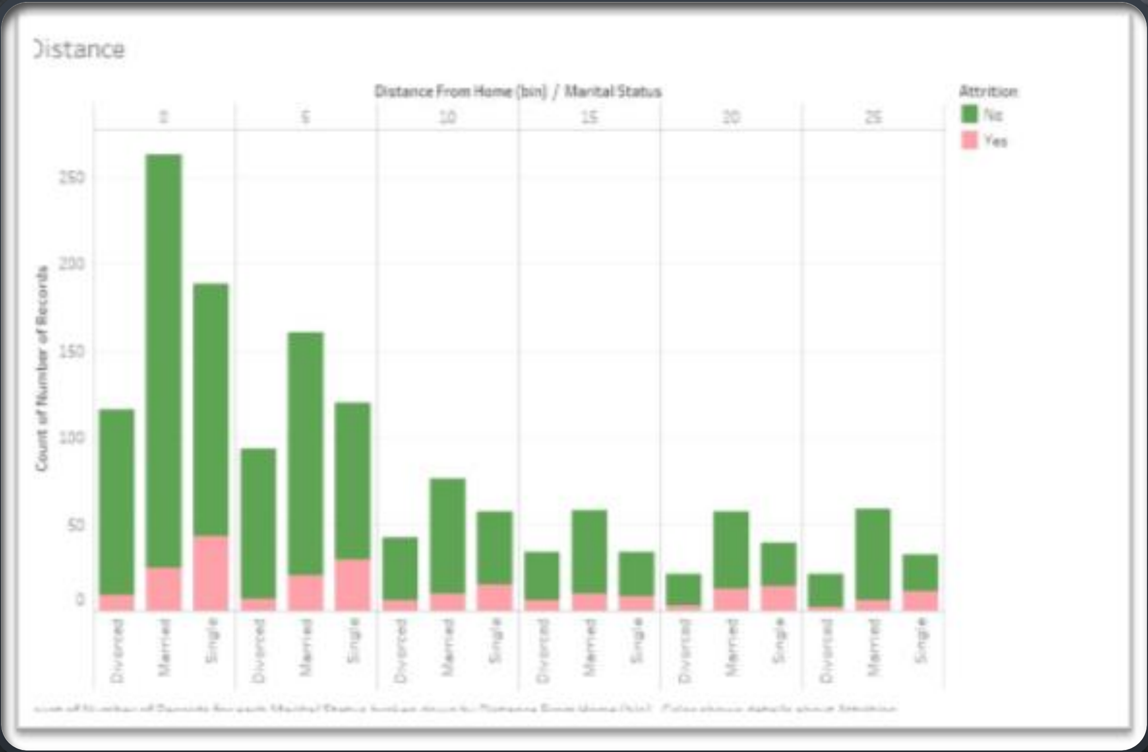
* The Smart framework is from Fractal analytics

Fractal

HOW CAN WE REDUCE IBM COMPANY'S ATTRITION RATE BY PREDICTING IF A CANDIDATE WILL EXIT IN INDIA WITHIN THE YEAR

1)Background	**4)Constraints**
• Best performing IT company in terms of topline and bottom line. • Employee Attrition has increased from 14% to 25% in the last 1 year, much higher than the industry average of 16% • Committed to curtailing attrition as it is not sustainable in the long run	• Salary hike cannot be considered
2)Desired Outcome	**5)Stakeholders**
• Reduce attrition rate by 5% in the next 18 months • Saving recruitment cost and improving employee satisfaction rate • develop a holistic employee satisfaction program	• CEO/ HR Head/ BU Heads • Attrition cell
3)Scope	**6)Resources**
• In-house attrition analysis tool and Early Warning System to identify individuals who are likely to leave and prioritize action items for immediate intervention	• Interviews with HR head, Attrition cell, recruitment team, BU Heads • Insights based on industry best practices and secondary research • Review of exit interviews and HR attrition data to observe trends across departments, gender, experience level , etc.

Exploratory Data Analysis



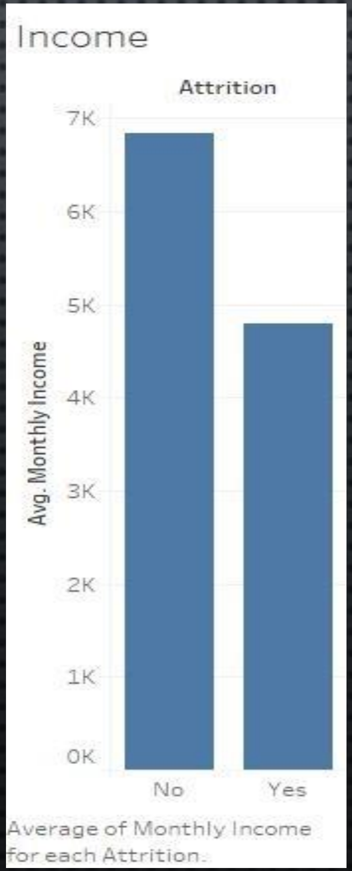
Our Aim is to analyse the IBM Attrition data and find what trends or patterns its showing towards attrition. To understand the data better, we deep dive into the data and try to find what could be the possible reason for the attrition (the analysis has been done step wise taking each variable).

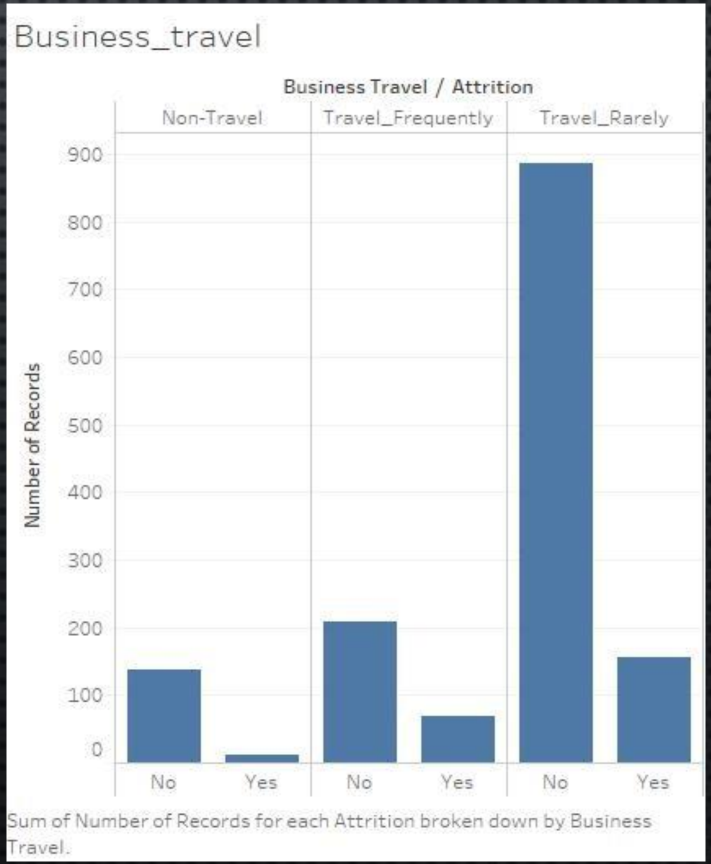
From our graphs we realized several interesting patterns, like that people of age between 25-35 and people who didn't get chance for business travel have high attrition rate, or that employees who are married and living at a distant place from company have high attrition rate, too.

Furthermore, we have been able to understand the principal categories with the highest attrition rates (i.e.: sales department, hr & technical degree in education, singles, males, low income, etc...)

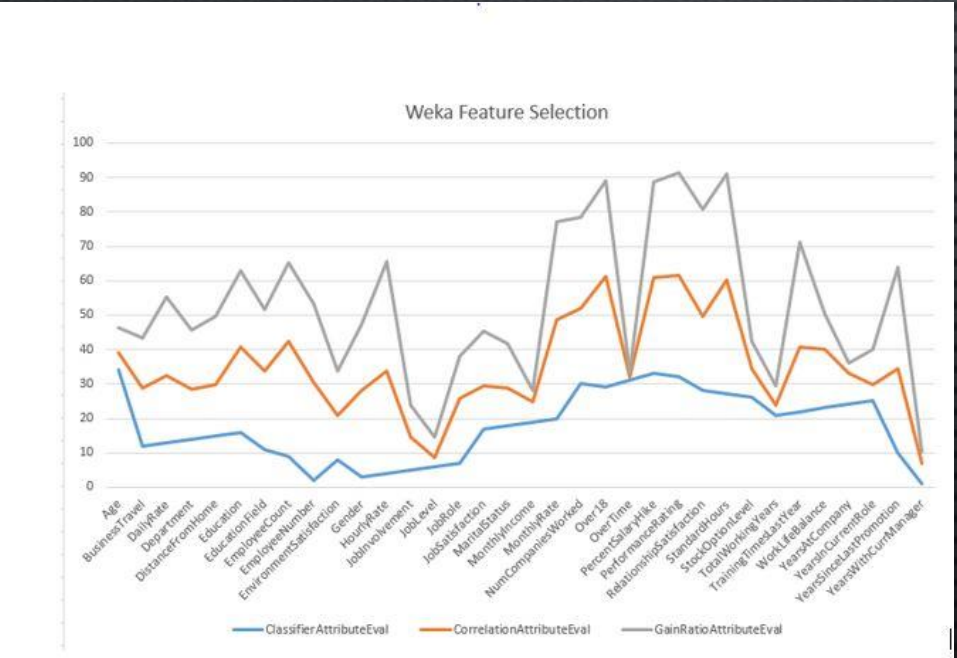
EXPLORATORY DATA ANALYSIS

- From the graph we observe that average monthly income affect attrition rate. Also people who do not get the chance to travel for business have higher attrition rates.





FEATURE SELECTION



Several techniques were used in Weka for feature selection

- Classifier Attribute Evaluation, Correlation Attribute Evaluation and Gain Ratio Attribute Evaluation. The chart below summarizes the results that these models found

FEATURE SELECTION

Note: No (additional) effects met the 0.05 significance level for removal from the model.

Summary of Backward Elimination					
Step	Effect Removed	DF	Number In	Wald Chi-Square	Pr > ChiSq
1	Education	4	29	0.7877	0.9401
2	PerformanceRating	1	28	0.0779	0.7801
3	Department	2	27	0.5313	0.7667
4	PercentSalaryHike	1	26	0.2384	0.6253
5	MonthlyRate	1	25	0.5345	0.4647
6	HourlyRate	1	24	0.6436	0.4224
7	MaritalStatus	2	23	2.2497	0.3247
8	MonthlyIncome	1	22	1.8706	0.1714
9	DailyRate	1	21	3.7932	0.0515

- Using the Backward elimination method, we were able to identify the following 9 variables that were not useful in predicting attrition in our classification model:

Data Cleaning and Balancing

- Well our data is serious imbalanced, so the first challenge we faced was balancing it (83/16)%
- Second we had a lot of factor variables which we needed to convert to dummies before beginning
- Except this our data was squeaky clean

CLASSIFICATION & INTERPRETATION

In order to create an early warning system we created a machine learning model that would perform classification to check if an employee could quit the company

Model	Data	ACC	ROC	Kappa
Logistic Regression	All data	88.04%	0.67	0.42
	Smote 1	81.52%	0.82	0.63
	Smote 2	79.25	0.79	0.58
	Fs via Backward elimination	86.41%	0.64	0.34
	Fs via CFS subset evaluation	85.32%	0.52	0.05
Decision Tree	All Data	84.23%	0.61	0.25
	Smote 1	85.57%	0.86	0.71
	Smote 2	92.22%	0.92	0.84
	Fs via Backward elimination	83.96%	0.64	0.31
	Fs via CFS subset evaluation	84.23%	0.61	0.25
Random Forest	All data	87.72%	0.59	0.26
ANN	All data	85.86%	0.64	0.33
SVM	All data	83.9%	0.58	0.19
XGBOOST	All data	88.04%	0.66	0.4
	Smote 1	91.41%	0.92	0.82
	Smote 2	87.35%	0.87	0.74

Classification & Interpretation

From the previous results its clear that decision tree stole the show!

However lets think practically

- It is often required to explain the business why we think a person could leave, in this case we need a model whose output we can explain. In our case a decision tree or logistic regression
- Sometimes HR would just like to run our model on random data sets , so its not always possible to Balance our datasets using techniques like smote
- Our model should just be able to predict better than random but imagine the cost of entertaining an employee who was not going to leave but our system tagged him – This is a future improvement for our model
- XGBoost model created a nice ensemble of trees for us, whose accuracy could increase more than the decision tree if we get more data

Clustering & Interpretation

We served ourselves of cluster techniques but for classification purposes. We used the K-Means method because it allows us to predetermine the number of output clusters. Since we are studying attrition, that in the dataset is a discrete binary class, we tuned the algorithm to get just two clusters.

We then use the train-test set splitting and applied the clustering on the training set. After that we evaluated both the validity of the clustering output and the classifying power of the same, through a set of measures (Distortion, Silhouette, Precision and Recall, ROC Area Under the Curve)

Going through, we reiterate the same exact process but with two other different datasets, obtained with different Feature Selection techniques. The first one was obtained with a Backward Elimination based on a Logistic regression model and run on SAS, the second one was obtained with a Correlation-based Feature Selection (CFS) run on Weka.

- After that we collected the result from every attempt and compared them, and as a result the most efficient with respect to the classification goal was the Backward Elimination that had a 0.78 precision.

ASSOCIATION RULE MINING

- Attrition=No JobRole=Sales Executive 269 ==> Department=Sales 269 <conf:(1)> lift:(3.3) lev:(0.13) [187] conv:(187.39)
- Attrition=No JobRole=Research Scientist 245 ==> Department=Research & Development 245 <conf:(1)> lift:(1.53) lev:(0.06) [84] conv:(84.83)
- Attrition=No JobRole=Research Scientist MonthlyIncome=Poor 244 ==> Department=Research & Development 244 <conf:(1)> lift:(1.53) lev:(0.06) [84] conv:(84.49)

CONCLUSIONS

- We successfully created an early warning system which immediately tells the Human Resources department if an employee is prune to leave or not.
- We achieved this early warning system based on several data mining techniques in order to be very accurate on supervised classification modelling

PERSONAL EVALUATIONS

TASK	RESPONSIBLE	REVIEWER
Performance Evaluation	Mohammed T.	Patricia L.
EDA	Igor P. Valerio T.	Millicent O.
Data Cleaning	Mohammed T.	Patricia L.
Features Selection	Particia L. Sanchita	Igor P. Millicent O.
Classification	Mohammed T. Patricia L.	Valerio T.
Clustering	Valerio T. Millicent O.	Mohammed T.
Interpretation	Mohammed T. Valerio T. Patricia L.	Igor P.
Report Making & PPT	Igor P. Sanchita Millicent O.	Mohammed T. Valerio T. Patricia L.

ISSUE TREE

how can we reduce ABC it company's attrition rate from 25% to 14% in India within the next 18 months

voluntary attrition

A1

employee satisfied with his / her job

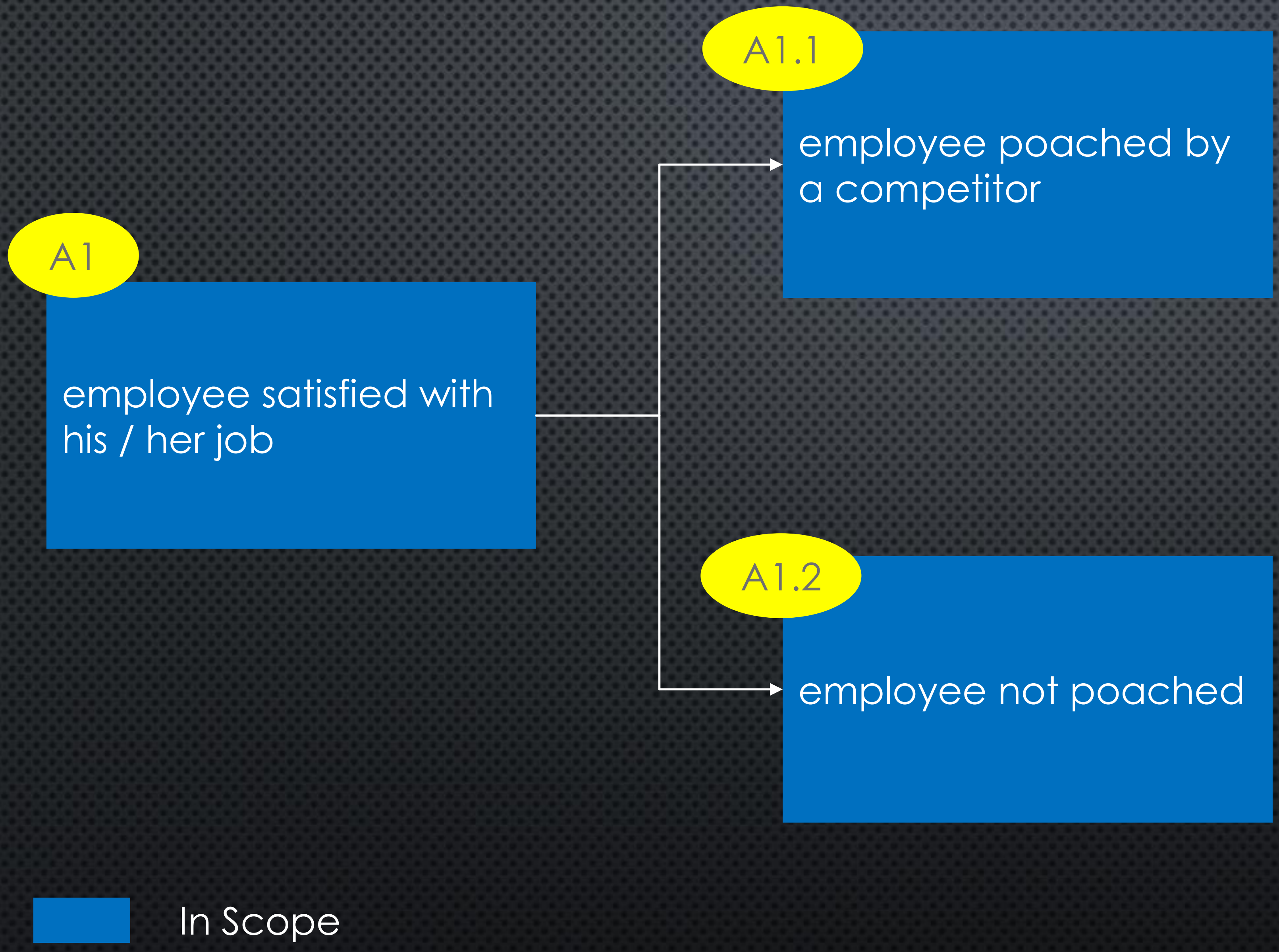
A2

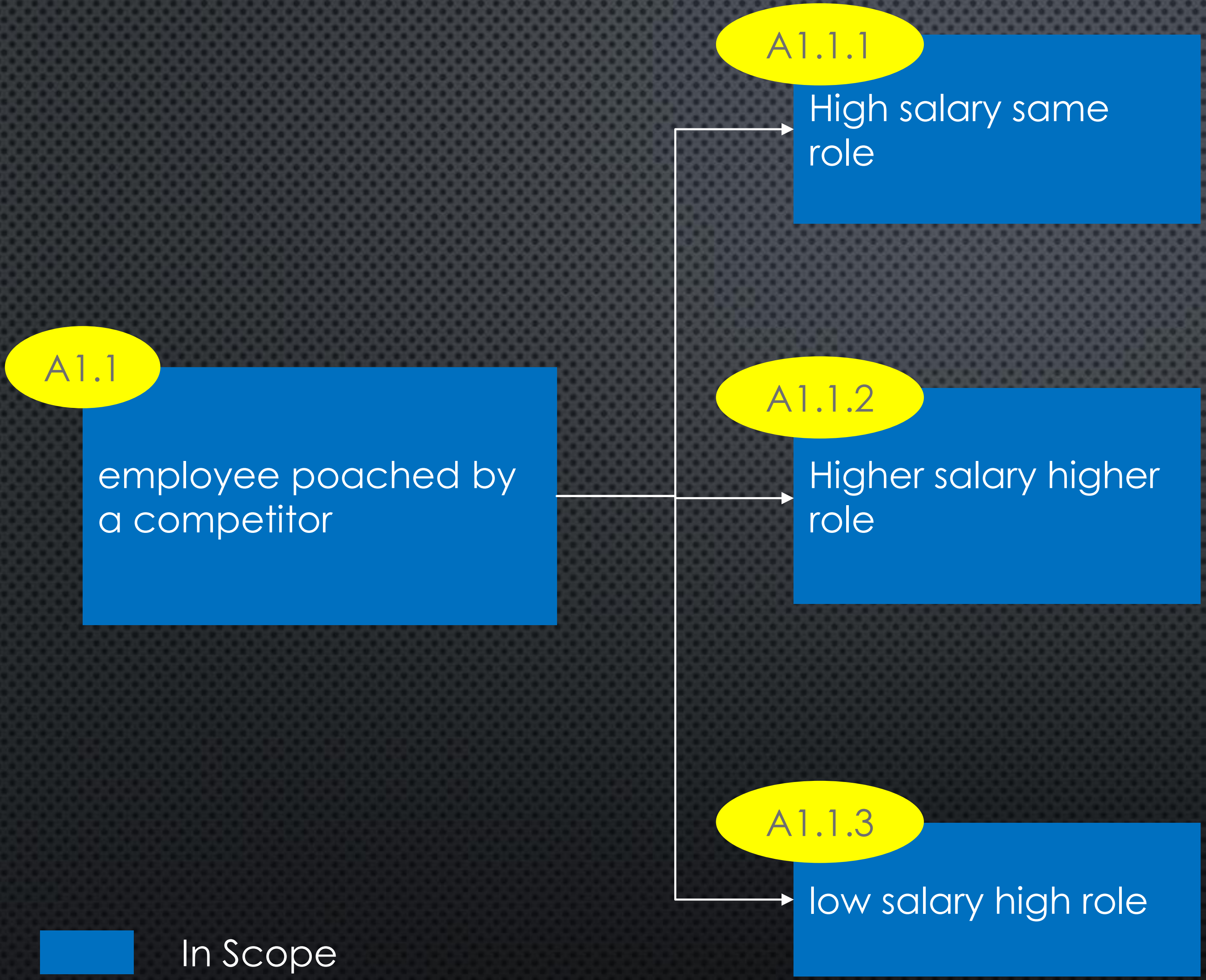
employee un-satisfied with his / her job

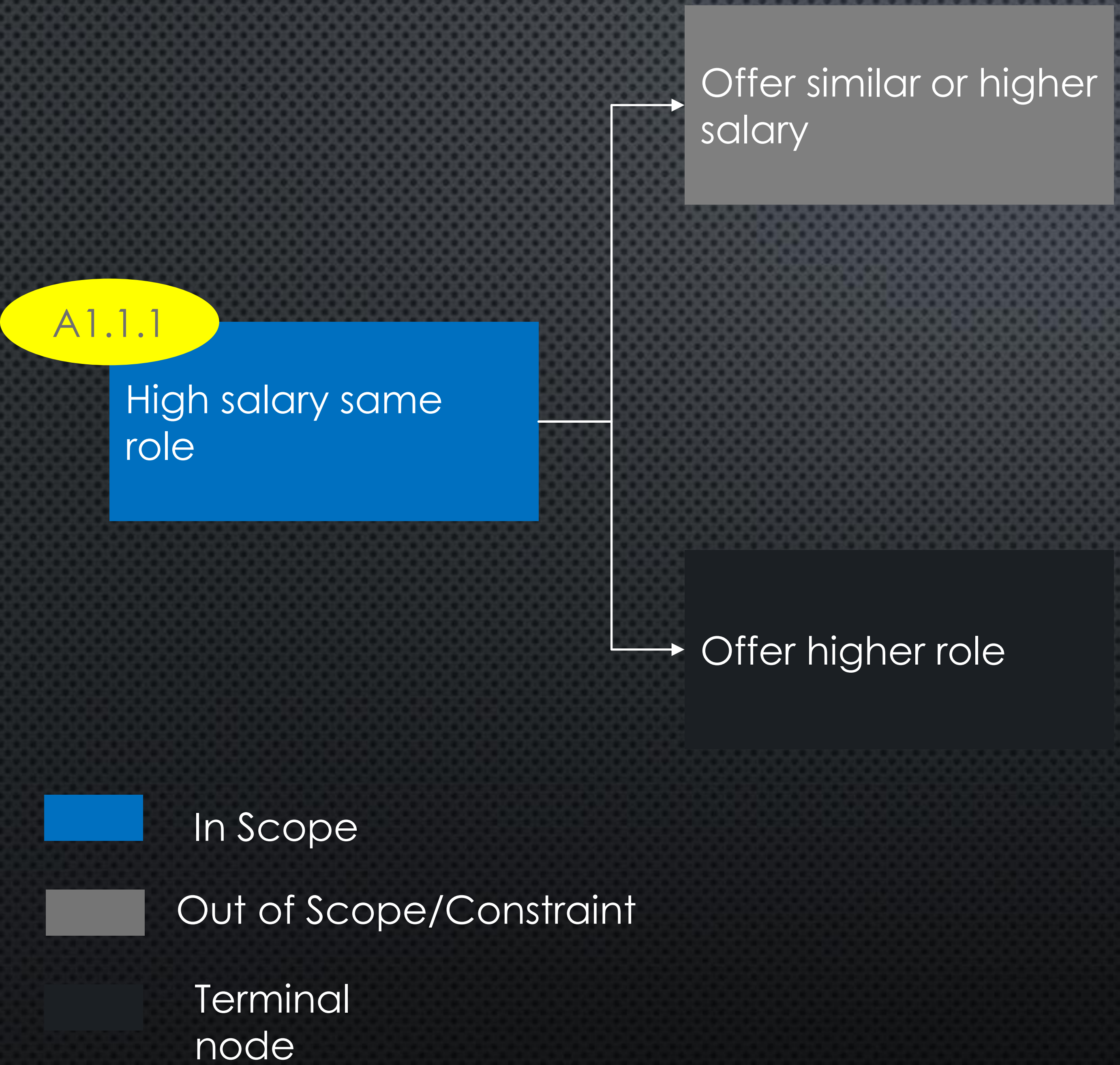
involuntary attrition i.e. employee has been asked to leave the organization

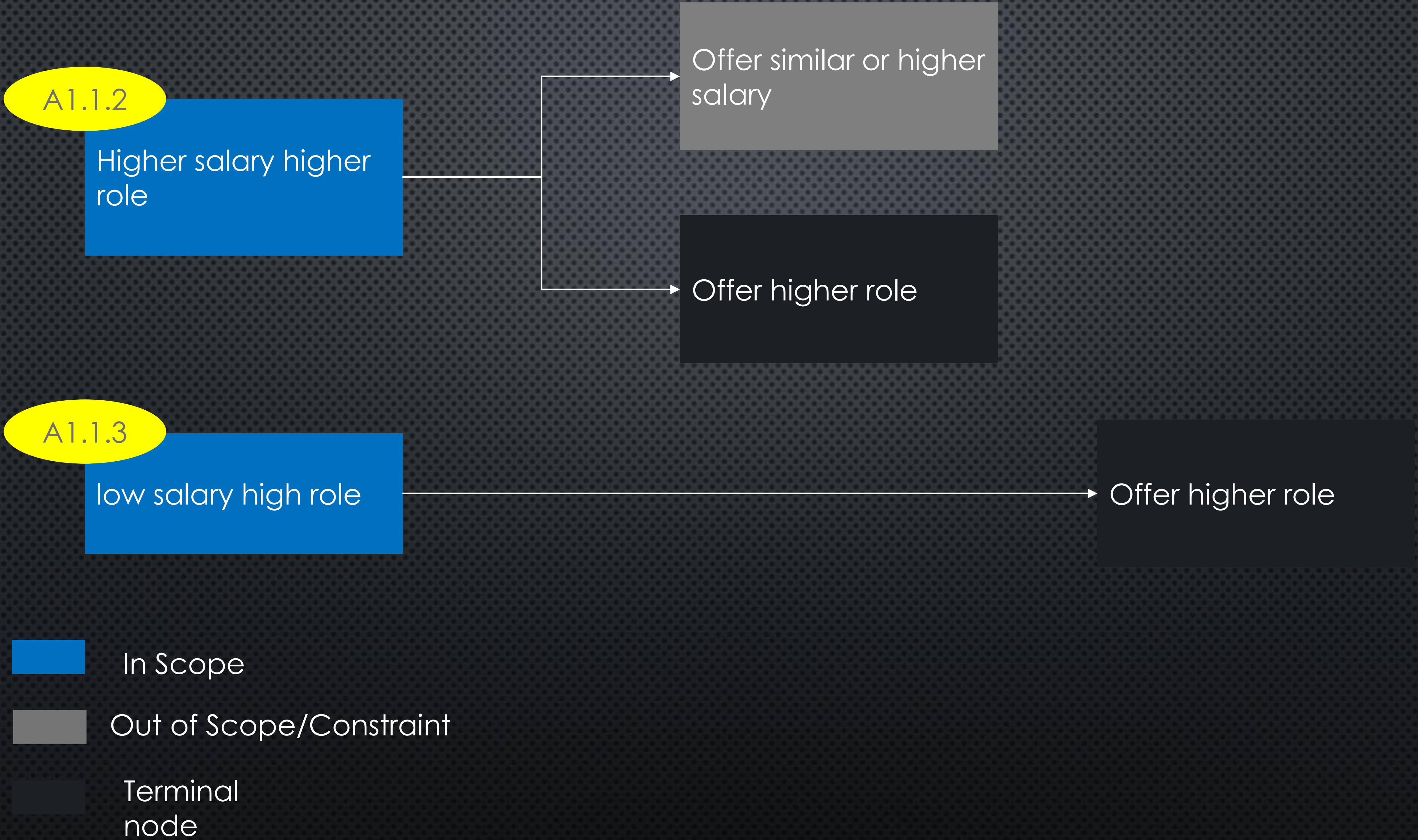
in scope

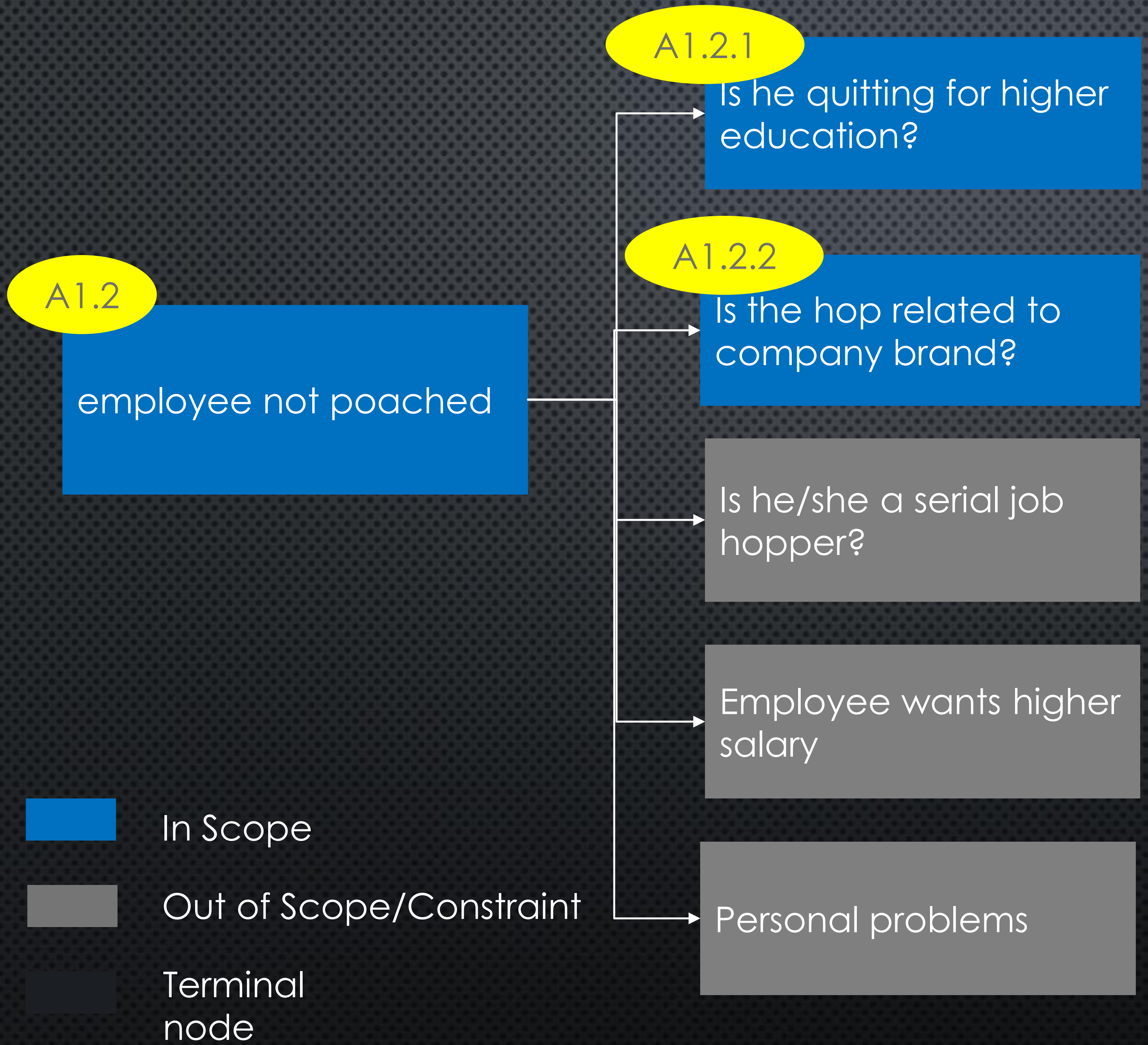
out of scope/constraint

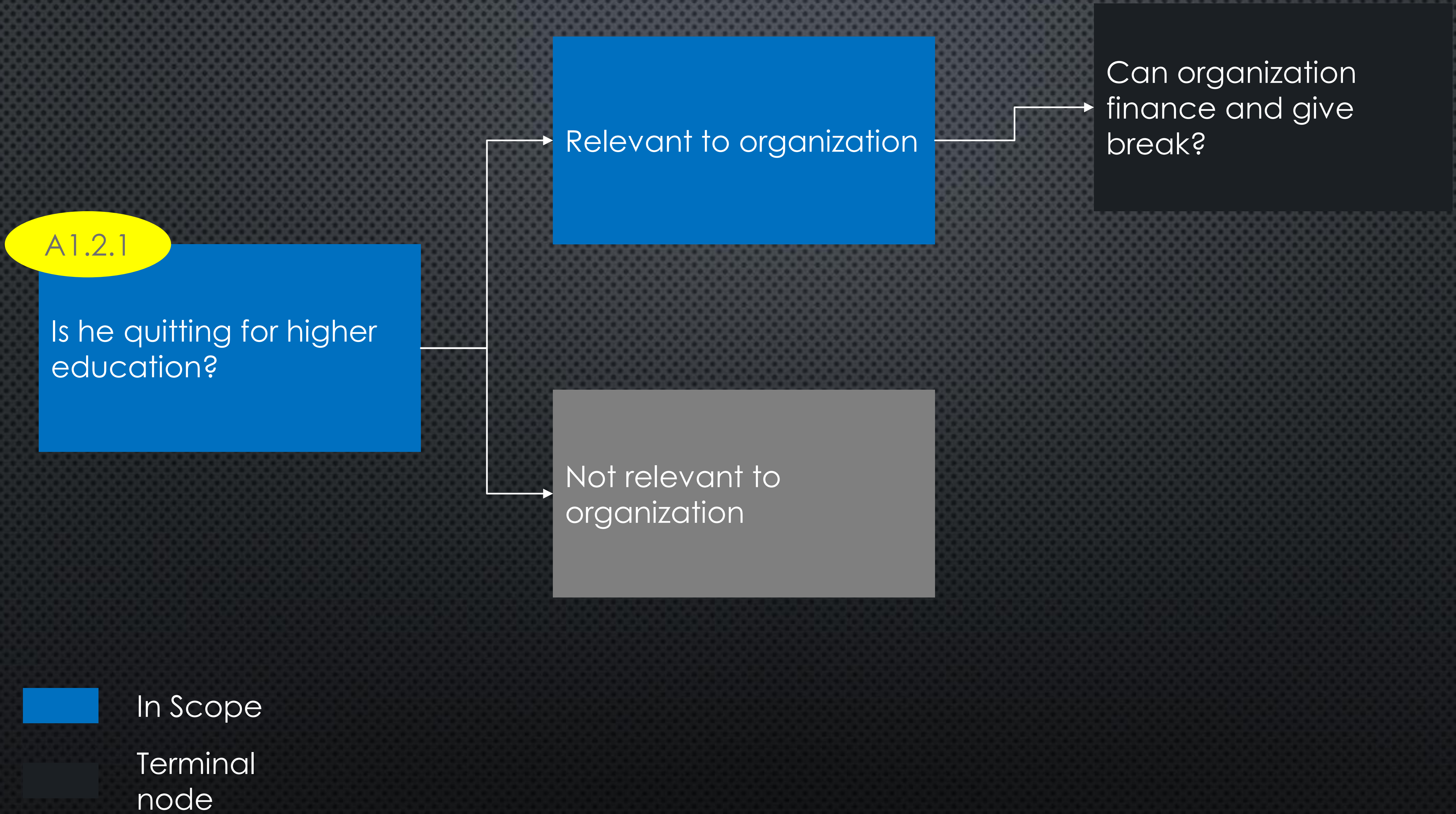

A1
employee satisfied with his / her job
A1.1
employee poached by a competitor
A1.2
employee not poached
In Scope

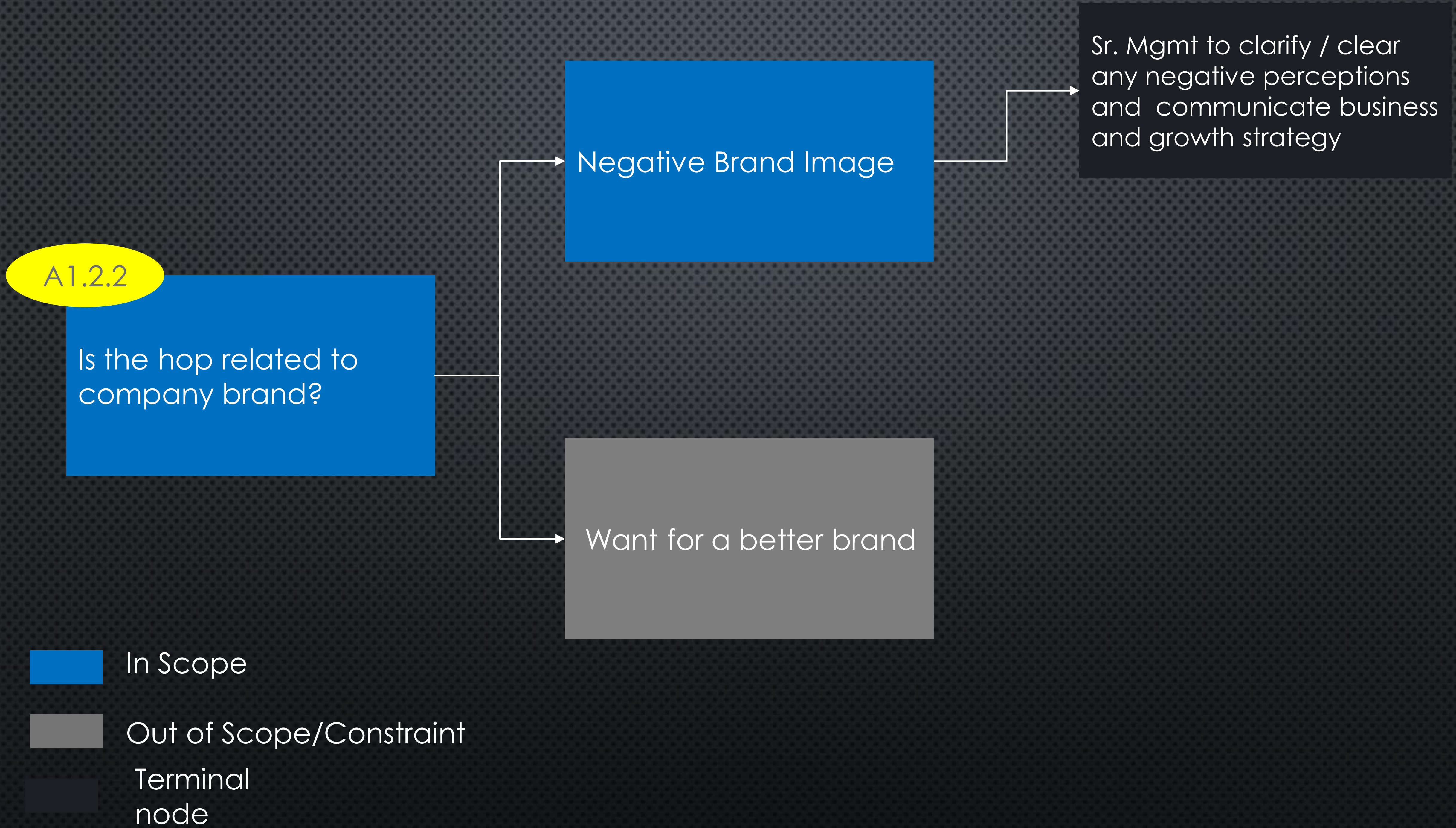

A1.1
employee poached by a competitor
A1.1.1
High salary same role
A1.1.2
Higher salary higher role
A1.1.3
low salary high role
In Scope

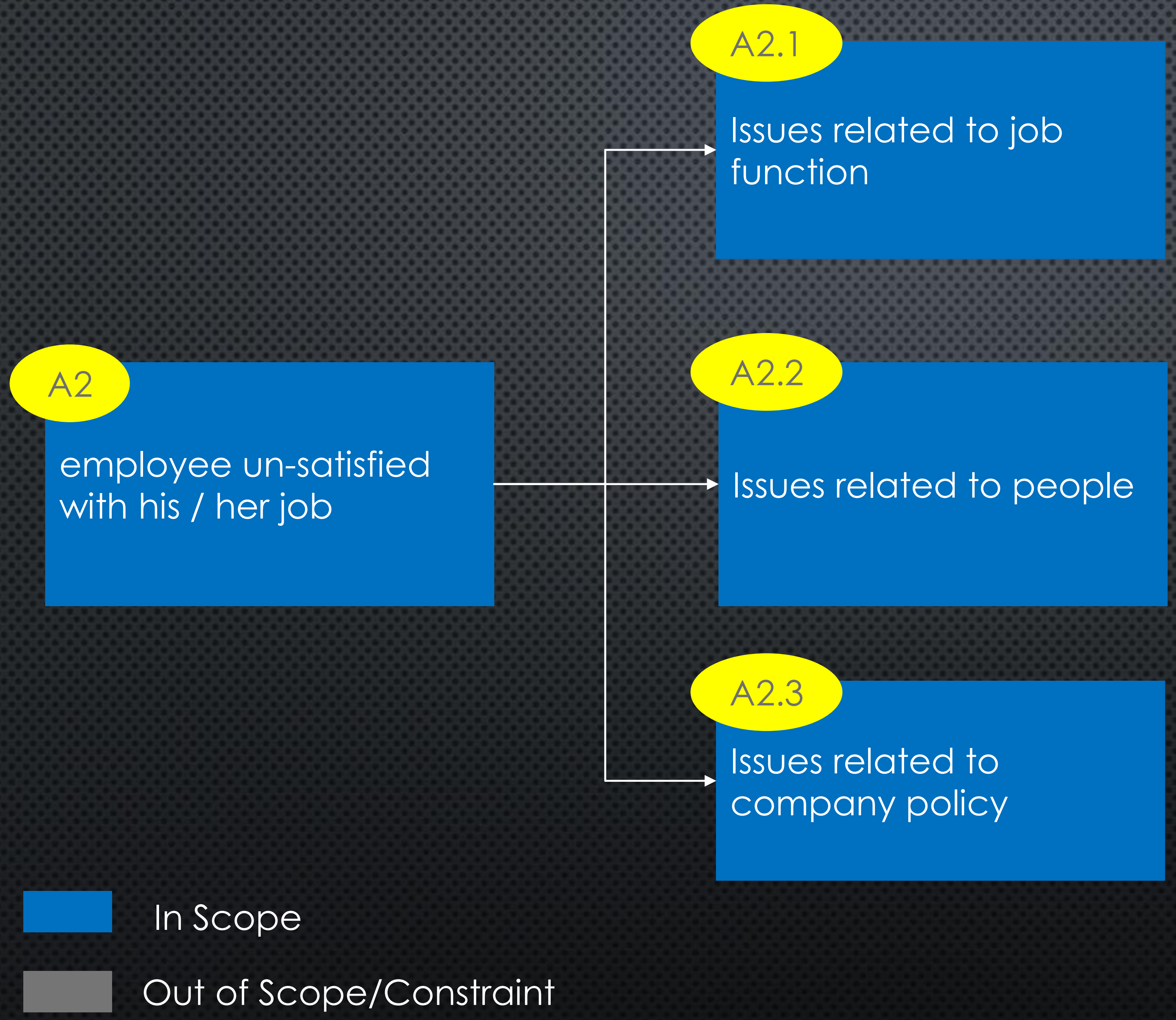

A1.1.1
High salary same role
Offer similar or higher salary
Offer higher role
In Scope
Out of Scope/Constraint
Terminal node

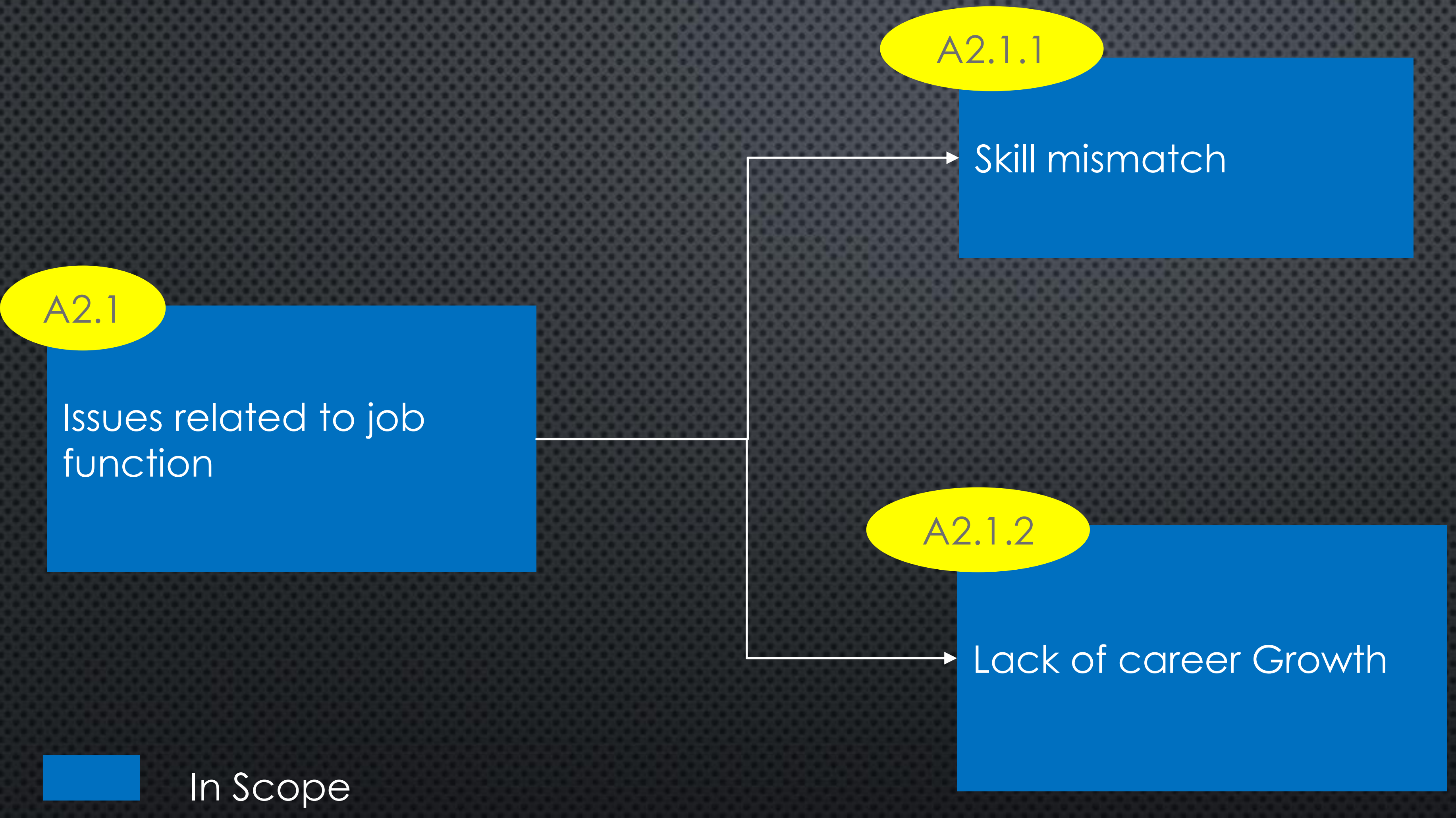

A1.1.2
Higher salary higher role
Offer similar or higher salary
Offer higher role
A1.1.3
low salary high role
Offer higher role
In Scope
Out of Scope/Constraint
Terminal node

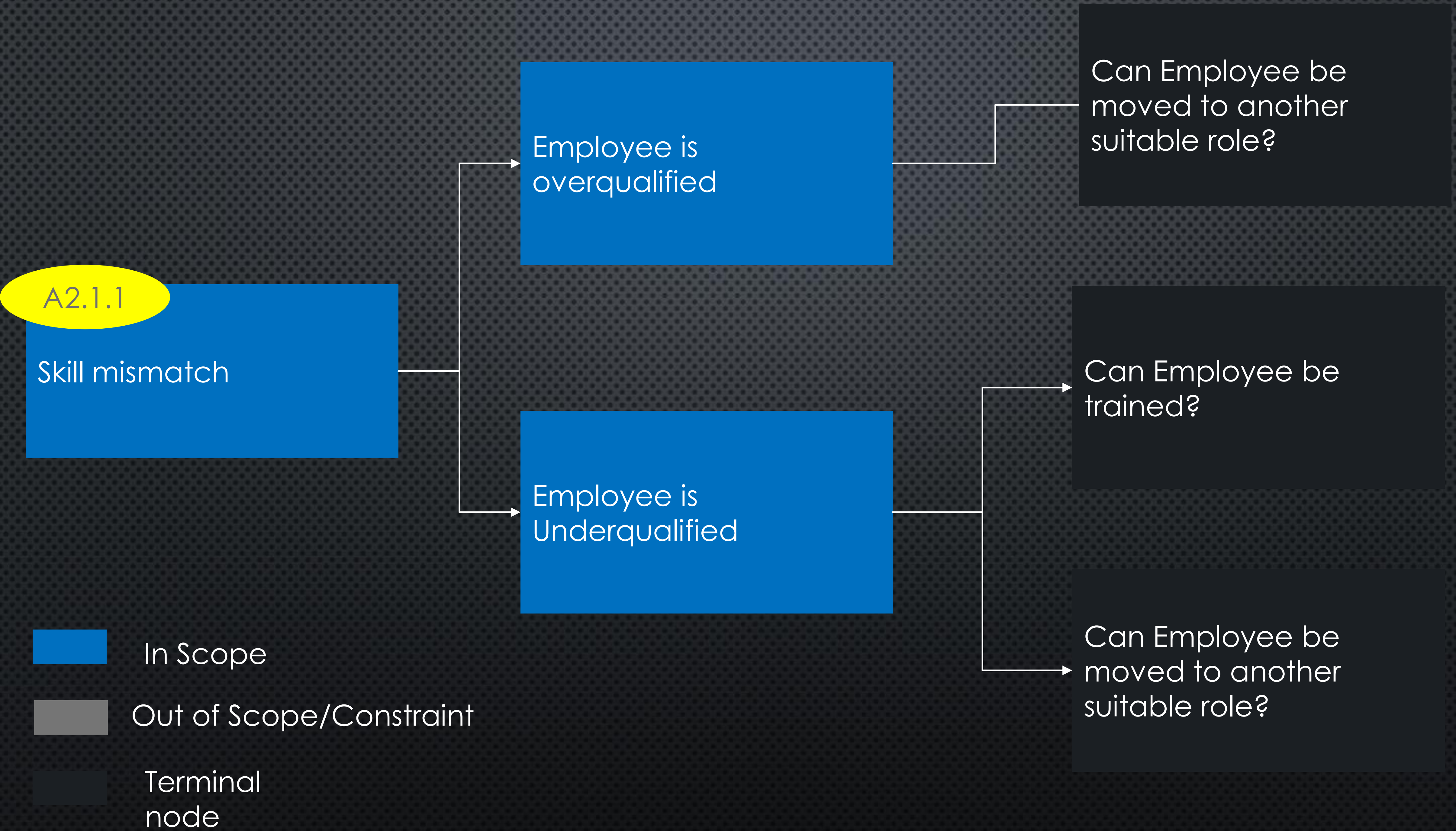

A1.2
employee not poached
A1.2.1
Is he quitting for higher education?
A1.2.2
Is the hop related to company brand?
Is he/she a serial job hopper?
Employee wants higher salary
Personal problems
In Scope
Out of Scope/Constraint
Terminal node

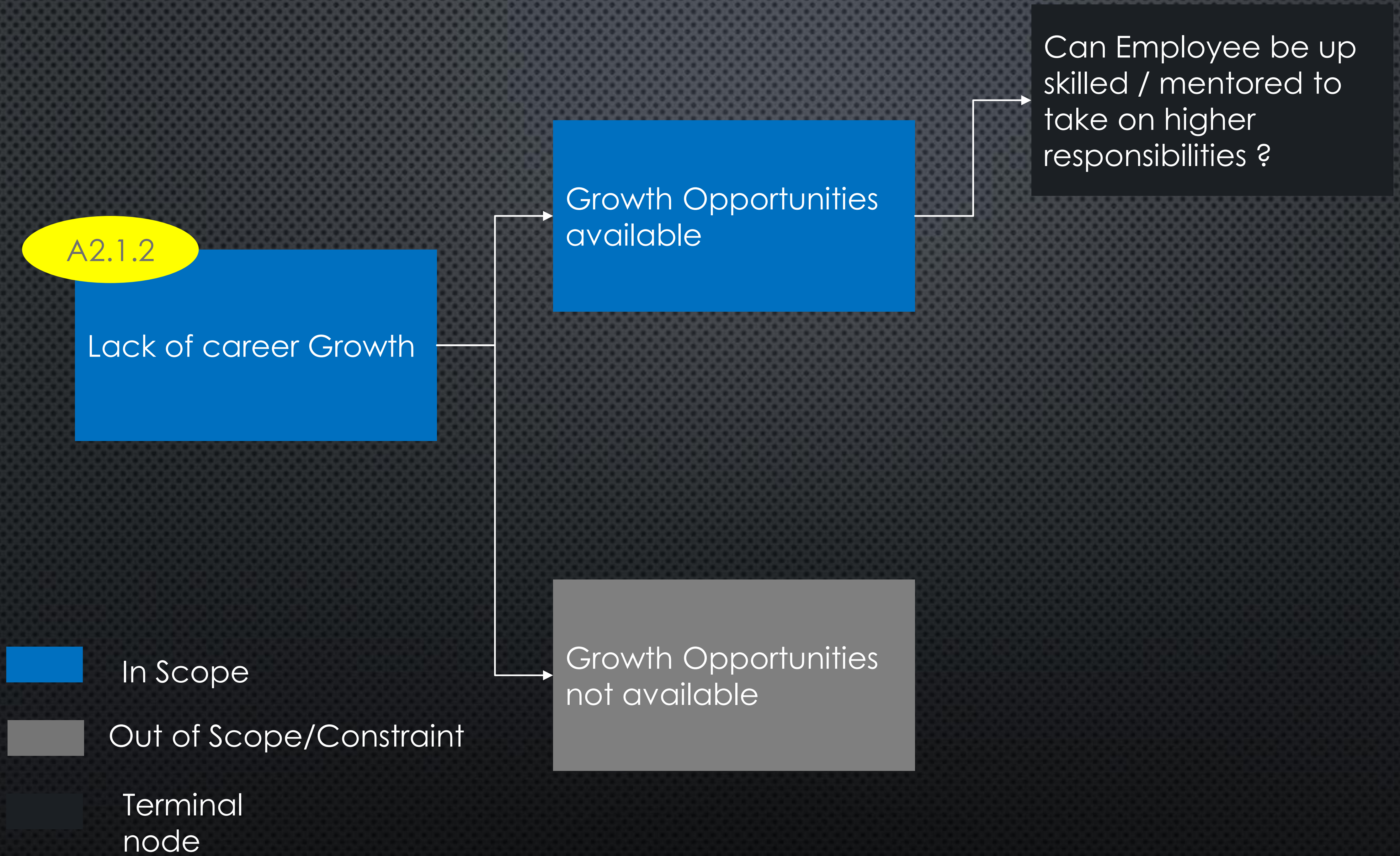

A1.2.1
Is he quitting for higher education?
Relevant to organization
Can organization finance and give break?
Not relevant to organization
In Scope
Terminal node

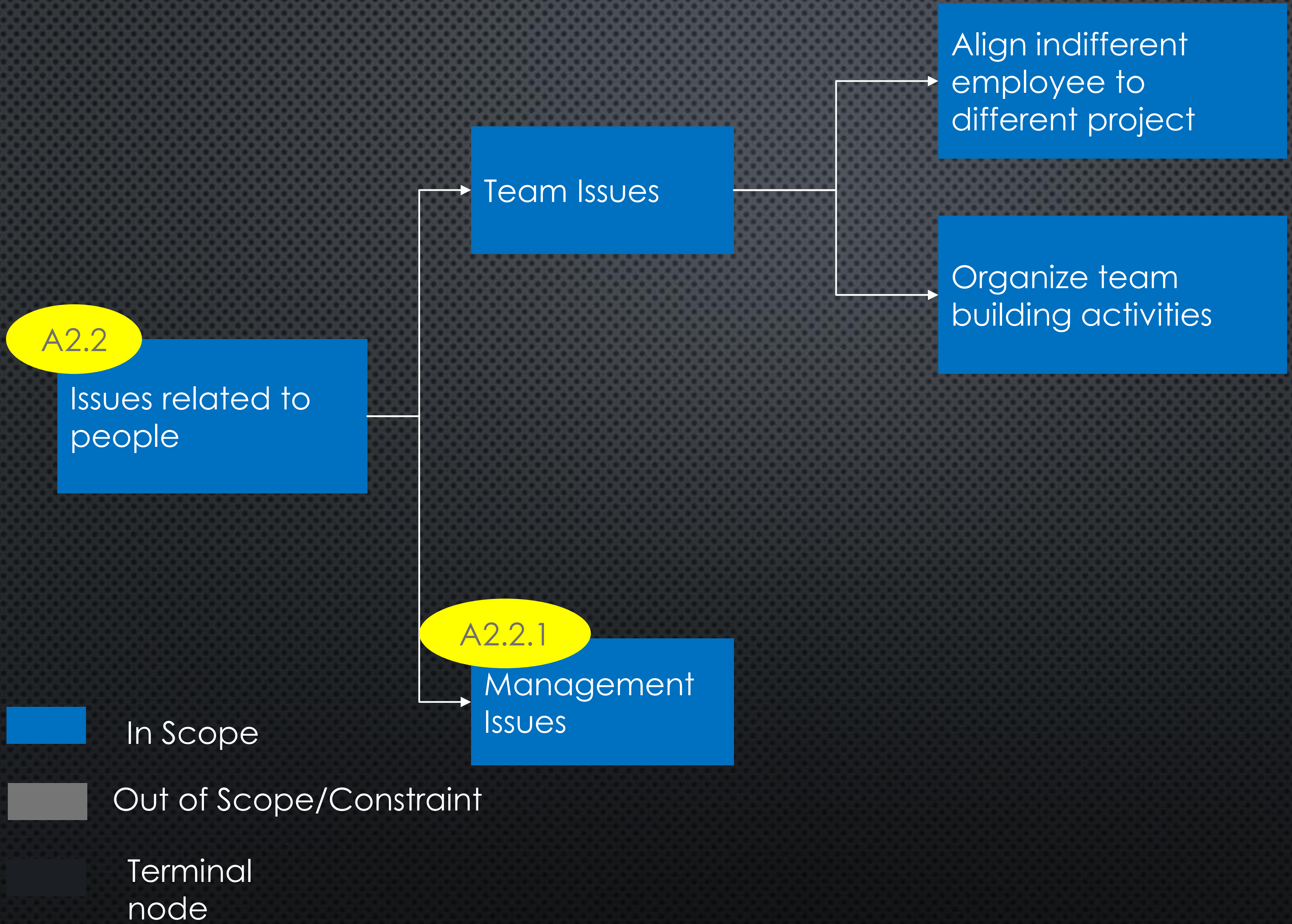

A1.2.2
Is the hop related to company brand?
Negative Brand Image
Want for a better brand
Sr. Mgmt to clarify / clear any negative perceptions and communicate business and growth strategy
In Scope
Out of Scope/Constraint
Terminal node

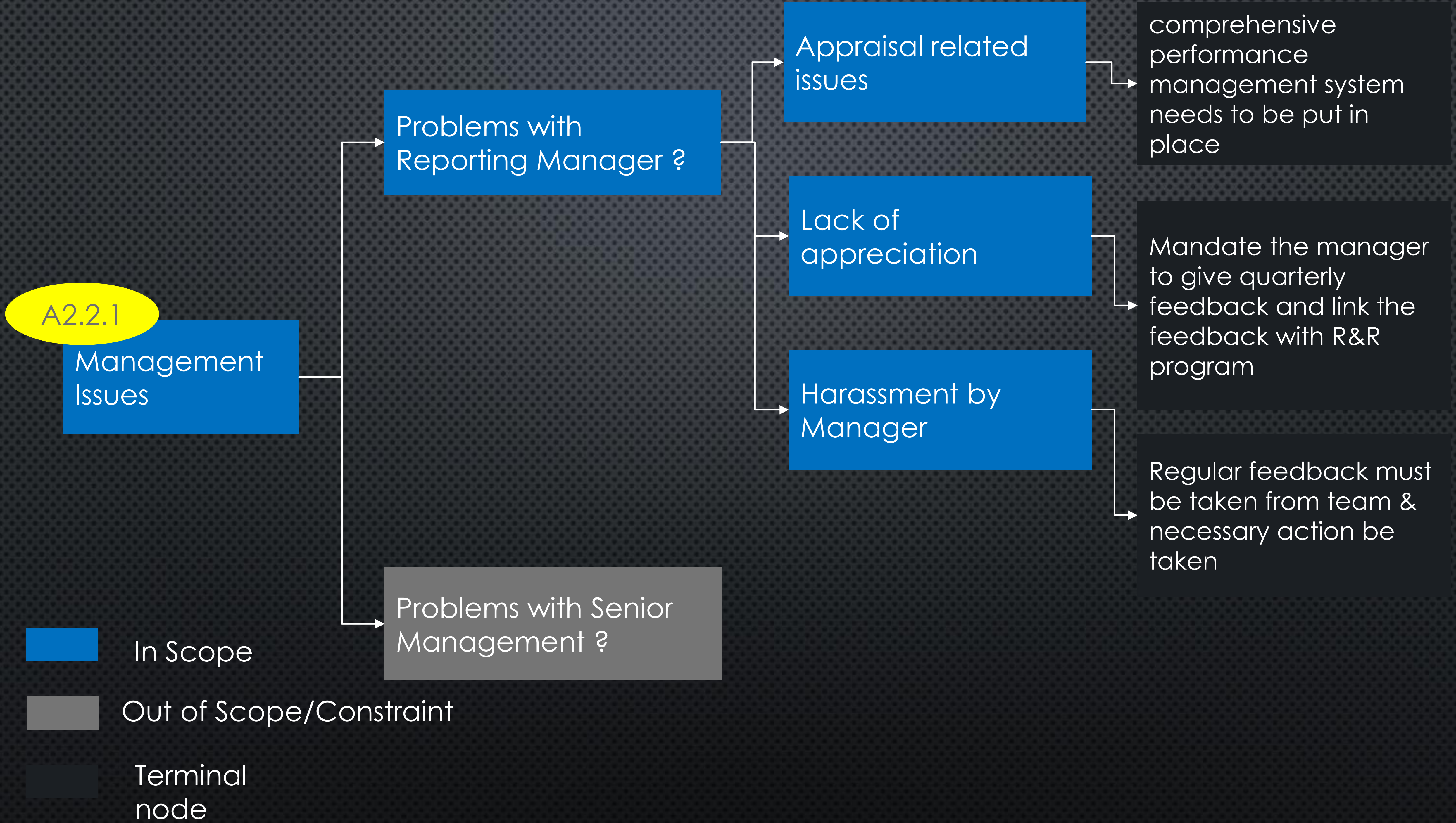

A2
employee un-satisfied with his / her job
A2.1
Issues related to job function
A2.2
Issues related to people
A2.3
Issues related to company policy
In Scope
Out of Scope/Constraint

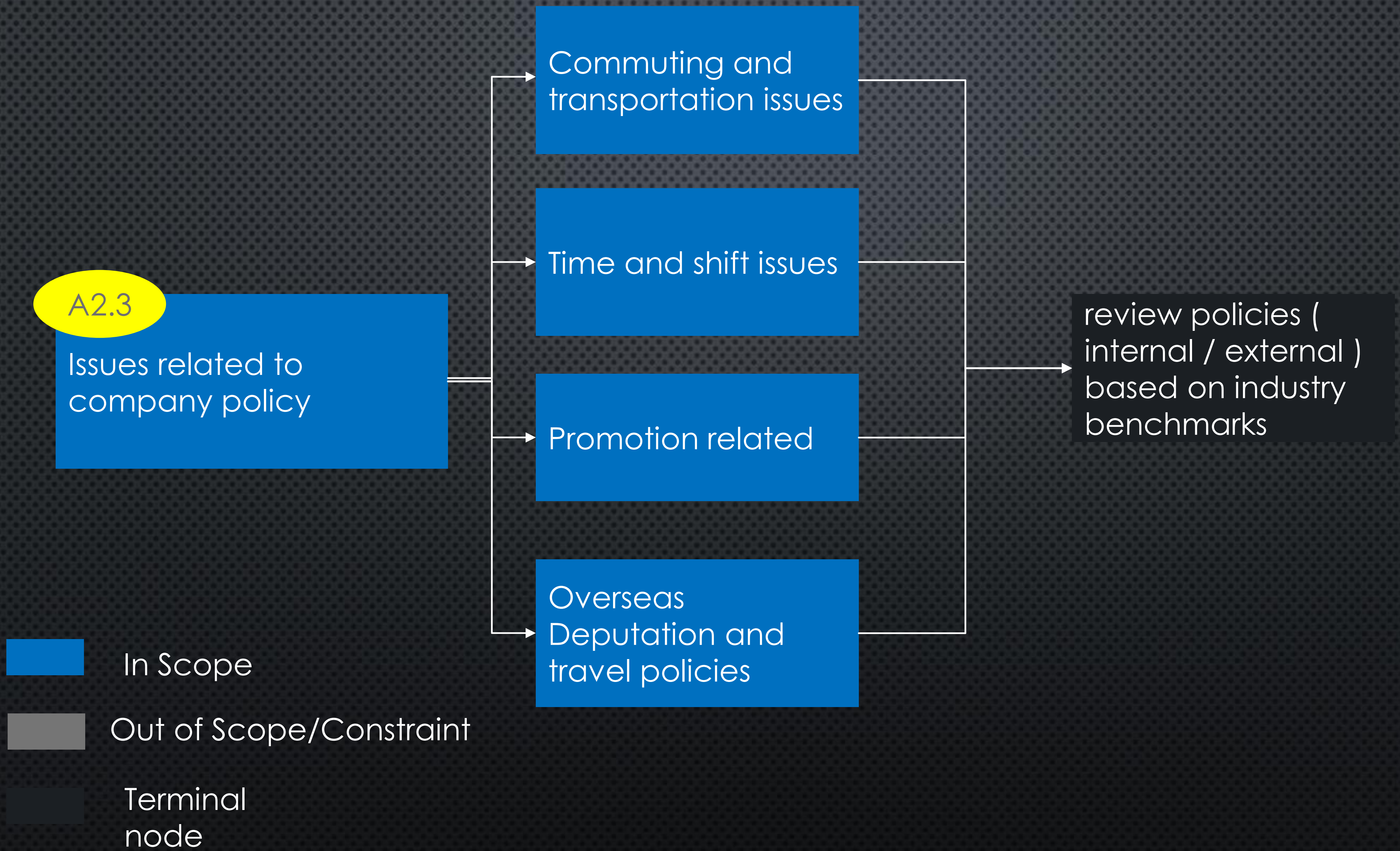

A2.1.1
Skill mismatch
A2.1
Issues related to job function
A2.1.2
Lack of career Growth
In Scope


A2.1.1
Skill mismatch
Employee is overqualified
Employee is Underqualified
Can Employee be moved to another suitable role?
Can Employee be trained?
Can Employee be moved to another suitable role?
In Scope
Out of Scope/Constraint
Terminal node


A2.1.2
Lack of career Growth
Growth Opportunities available
Can Employee be up skilled / mentored to take on higher responsibilities ?
Growth Opportunities not available
In Scope
Out of Scope/Constraint
Terminal node


A2.2
Issues related to people
Team Issues
Align indifferent employee to different project
Organize team building activities
A2.2.1
Management Issues
In Scope
Out of Scope/Constraint
Terminal node


A2.2.1
Management Issues
Problems with Reporting Manager ?
Problems with Senior Management ?
Appraisal related issues
Lack of appreciation
Harassment by Manager
comprehensive performance management system needs to be put in place
Mandate the manager to give quarterly feedback and link the feedback with R&R program
Regular feedback must be taken from team & necessary action be taken
In Scope
Out of Scope/Constraint
Terminal node


A2.3
Issues related to company policy
Commuting and transportation issues
Time and shift issues
Promotion related
Overseas Deputation and travel policies
review policies (internal / external) based on industry benchmarks
In Scope
Out of Scope/Constraint
Terminal node

EARLY WARNING SYSTEM - MODELLING