Weight Sum of Helpfulness Metrics

Here we are going to create a function that prints out the importance of each helpfulness metric. Once we have the function, it can be used to look at the overall weights as well as weights for individual workers.

In [3]:
```
import pandas as pd
import numpy as np
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import StandardScaler

def calculate_metric_weights(df, worker_id=None):
    '''
    Calculate the weight of each helpfulness metric for a given writer.

    Input: csv file

    Output: dictionary of feature importance
    '''

    # Select relevant columns
    features = ['tone_similarity',      'pos_similarity',       'coherence_sco
    target = 'acceptance_status'

    # Filter for specific writer if worker_id is provided
    if worker_id:
        df = df[df['workerID'] == worker_id]

    # Drop rows with missing values
    df = df.dropna(subset=features + [target])

    # Convert target to binary, accepted is 1 and rejected is 0
    df[target] = df[target].apply(lambda x: 1 if x == 'accepted' else 0)

    if df[target].nunique() < 2:
        print(f"WorkerID: {worker_id} has only one class in target variable. Sk
        return {}

    # Standardize the features
    scaler = StandardScaler()
    X = scaler.fit_transform(df[features])
    y = df[target]

    # Train logistic regression model
    model = LogisticRegression()
    model.fit(X, y)

    # Get absolute values of coefficients and normalize them to sum to 1
    weights = np.abs(model.coef_[0])
    weights /= weights.sum()

    # Return dictionary of feature importance
    return dict(zip(features, weights))
```

In [4]:
```
df = pd.read_csv("all_metrics_upd.csv")

# Get overall weights
weights = calculate_metric_weights(df)
print(weights)

# Get weights for all writers
unique_workers = df['workerID'].unique()

for worker in unique_workers:
  weights = calculate_metric_weights(df, worker_id=worker)
  print(f'WorkerID: {worker}, Weights: {weights}')
```

{'tone_similarity': 0.07758531261918378, 'pos_similarity': 0.3341039854969208, 'coherence_score': 0.45104751656498815, 'syn_similarity': 0.03702799519444506, 'ai_coverage': 0.10023519012446222}
WorkerID: A2WGW5Y3ZFBDEC, Weights: {'tone_similarity': 0.2320297946292556, 'pos_similarity': 0.1901218311849363, 'coherence_score': 0.011560217587770755, 'syn_similarity': 0.37453692461188215, 'ai_coverage': 0.19175123198615507}
WorkerID: AZCGF2D7QIO10, Weights: {'tone_similarity': 0.03929922174212228, 'pos_similarity': 0.26295291155230843, 'coherence_score': 0.2447686117926599, 'syn_similarity': 0.2838165481685913, 'ai_coverage': 0.1691627067443181}
WorkerID: A345TDMHP3DQ3G, Weights: {'tone_similarity': 0.06473990785379341, 'pos_similarity': 0.3060658322749765, 'coherence_score': 0.4475075526413485, 'syn_similarity': 0.14508921036295336, 'ai_coverage': 0.03659749686692819}
WorkerID: A2W121DQXNQK1, Weights: {'tone_similarity': 0.38613892230010555, 'pos_similarity': 0.038879412895732296, 'coherence_score': 0.4705321459642223, 'syn_similarity': 0.01147490240872356, 'ai_coverage': 0.09297461643121624}
WorkerID: A324VBRLXHG5IB, Weights: {'tone_similarity': 0.17525337182351772, 'pos_similarity': 0.38997681820877617, 'coherence_score': 0.18071719082171034, 'syn_similarity': 0.046529285691671585, 'ai_coverage': 0.20752333345432422}
WorkerID: A2ONILC0LZKG6Y, Weights: {'tone_similarity': 0.26428489911731967, 'pos_similarity': 0.016768660464536338, 'coherence_score': 0.24173929067772812, 'syn_similarity': 0.3423082927290458, 'ai_coverage': 0.13489885701137008}
WorkerID: A3OLRWACCCCUTU, Weights: {'tone_similarity': 0.0930884667537779, 'pos_similarity': 0.00629922539415446, 'coherence_score': 0.4455648081088858, 'syn_similarity': 0.21545738579782867, 'ai_coverage': 0.2395901139453533}
WorkerID: A2EED3HLTA96CP, Weights: {'tone_similarity': 0.17223924837851584, 'pos_similarity': 0.0598854945668866, 'coherence_score': 0.5330058741071557, 'syn_similarity': 0.13572662854511175, 'ai_coverage': 0.09914275440233007}
WorkerID: A2QKAA5YS0P4CI, Weights: {'tone_similarity': 0.20016460645426962, 'pos_similarity': 0.0743278541293868, 'coherence_score': 0.3358742859288678, 'syn_similarity': 0.15902291410968525, 'ai_coverage': 0.23061033937779055}
WorkerID: A3S67QA0SQVPUJ, Weights: {'tone_similarity': 0.4441463057905086, 'pos_similarity': 0.10710095336570472, 'coherence_score': 0.20613860416395152, 'syn_similarity': 0.12592529637788497, 'ai_coverage': 0.11668884030195016}
WorkerID: A1198W1SPF1R4, Weights: {'tone_similarity': 0.2567388092993149, 'pos_similarity': 0.19019628719844853, 'coherence_score': 0.18669884017244698, 'syn_similarity': 0.2252331771833431, 'ai_coverage': 0.14113288614644645}
WorkerID: ANCIB6B6EBBIJ, Weights: {'tone_similarity': 0.04630325308870976, 'pos_similarity': 0.12641813277816924, 'coherence_score': 0.34372094841648976, 'syn_similarity': 0.28722260923990706, 'ai_coverage': 0.19633505647672428}
WorkerID: AZZA3J049G7R5, Weights: {'tone_similarity': 0.17388101541846226, 'pos_similarity': 0.21198850412195802, 'coherence_score': 0.0889309531492502, 'syn_similarity': 0.25629260874955667, 'ai_coverage': 0.2689069185607728}
WorkerID: A8C3WNWRBWUXO, Weights: {'tone_similarity': 0.27546868464584356, 'pos_similarity': 0.035790552050490455, 'coherence_score': 0.359559337533846, 'syn_similarity': 0.2104186358487334, 'ai_coverage': 0.11876278992108658}
WorkerID: A1PTH9KTRO06EG, Weights: {'tone_similarity': 0.16683095921779093, 'pos_similarity': 0.12445920153789543, 'coherence_score': 0.05409584016752882, 'syn_similarity': 0.2664566824908577, 'ai_coverage': 0.38815731658592717}
WorkerID: A2QX3YJXAAHHVV, Weights: {'tone_similarity': 0.049856239534255185, 'pos_similarity': 0.33425120552997206, 'coherence_score': 0.23711675355140577, 'syn_similarity': 0.01995044726151113, 'ai_coverage': 0.3588253541228558}
WorkerID: A23KAJRDVCVGOE, Weights: {'tone_similarity': 0.4105418619034455, 'pos_similarity': 0.31348458980863825, 'coherence_score': 0.16696154286860151, 'syn_similarity': 0.08024994461967684, 'ai_coverage': 0.028762060799637904}
WorkerID: A1QUQ0TV9KVD4C, Weights: {'tone_similarity': 0.07035202945253555, 'pos_similarity': 0.3497958321517789, 'coherence_score': 0.3355141115846414, 'syn_similarity': 0.14033623882144694, 'ai_coverage': 0.1040017879895972}
WorkerID: A1VZSFHTU51JP0, Weights: {'tone_similarity': 0.2990378986827508, 'pos_similarity': 0.20608443335317722, 'coherence_score': 0.3134213504531066, 'syn_similarity': 0.09745707108637669, 'ai_coverage': 0.08399924642458853}

```
WorkerID: A3MYPYBVHX7FQ2, Weights: {'tone_similarity': 0.011661566090877722,
'pos_similarity': 0.3159945765345165, 'coherence_score': 0.2801621723332339,
'syn_similarity': 0.30942474920127955, 'ai_coverage': 0.08275693584009225}
WorkerID: ASVRLMDNQBUD9, Weights: {'tone_similarity': 0.08374780942191837, 'po
s_similarity': 0.6278934676900523, 'coherence_score': 0.07233224204878734, 'sy
n_similarity': 0.17210640291708418, 'ai_coverage': 0.04392007792215794}
WorkerID: A2YTQDLACTLIBA, Weights: {'tone_similarity': 0.06862393645327286, 'p
os_similarity': 0.0061471762990999166, 'coherence_score': 0.2908990236669466,
'syn_similarity': 0.08629662850111278, 'ai_coverage': 0.5480332350795679}
WorkerID: A1TW2BZRRS874Z, Weights: {'tone_similarity': 0.11703882950673145, 'p
os_similarity': 0.29141156688559683, 'coherence_score': 0.34636112070377467,
'syn_similarity': 0.17897880901355237, 'ai_coverage': 0.06620967389034481}
WorkerID: A2OVX9UW5WANQE, Weights: {'tone_similarity': 0.2587646961326026, 'po
s_similarity': 0.07660101305434074, 'coherence_score': 0.3103998396567236, 'sy
n_similarity': 0.09646423498855133, 'ai_coverage': 0.25777021616778173}
WorkerID: A2OVOVZBJYUO, Weights: {'tone_similarity': 0.01886120234801627, 'pos
_similarity': 0.149843685979566, 'coherence_score': 0.45889645125144446, 'syn_
similarity': 0.1266921172495576, 'ai_coverage': 0.24570654317141583}
WorkerID: A1FVXS8IM5QYO8, Weights: {'tone_similarity': 0.18568723111875354, 'p
os_similarity': 0.2896271694285354, 'coherence_score': 0.3154996972200082, 'sy
n_similarity': 0.11038279180184861, 'ai_coverage': 0.0988031104308542}
WorkerID: A3O5RKGH6VB19C, Weights: {'tone_similarity': 0.013278031594197019,
'pos_similarity': 0.45114305937779364, 'coherence_score': 0.17545363121376864,
'syn_similarity': 0.23131083086446938, 'ai_coverage': 0.12881444694977123}
WorkerID: ABL2FXYMI00T6, Weights: {'tone_similarity': 0.2641982487526982, 'pos
_similarity': 0.09309929107467038, 'coherence_score': 0.3063927132935032, 'syn
_similarity': 0.2951790404523611, 'ai_coverage': 0.041130706426766965}
WorkerID: AZLZA0Q87TJZO, Weights: {'tone_similarity': 0.1745908287793147, 'pos
_similarity': 0.012234148116835696, 'coherence_score': 0.47221137792589707, 's
yn_similarity': 0.05117835469803314, 'ai_coverage': 0.28978529047991947}
WorkerID: A3P9TM5PRYBH90, Weights: {'tone_similarity': 0.02430096917646852, 'p
os_similarity': 0.14091726760815032, 'coherence_score': 0.33666603460671835,
'syn_similarity': 0.003631651075110871, 'ai_coverage': 0.494484077533552}
WorkerID: A118BQHK3S4UDV, Weights: {'tone_similarity': 0.32650524193876546, 'p
os_similarity': 0.05119434433461149, 'coherence_score': 0.33620527537439937,
'syn_similarity': 0.20561036406215083, 'ai_coverage': 0.08048477429007282}
WorkerID: A17Q4QN6UE0EZC has only one class in target variable. Skipping...
WorkerID: A17Q4QN6UE0EZC, Weights: {}
WorkerID: AFIK3VBMMX6G6 has only one class in target variable. Skipping...
WorkerID: AFIK3VBMMX6G6, Weights: {}
WorkerID: A14OQ52EFQAN2W, Weights: {'tone_similarity': 0.17977353838410748, 'p
os_similarity': 0.34838109596527883, 'coherence_score': 0.23559325065774486,
'syn_similarity': 0.014230335590505464, 'ai_coverage': 0.22202177940236345}
WorkerID: A1WKF2VH7TV0H2 has only one class in target variable. Skipping...
WorkerID: A1WKF2VH7TV0H2, Weights: {}
WorkerID: A3DUPRZSMU9W5R, Weights: {'tone_similarity': 0.02350892209587476, 'p
os_similarity': 0.09976606939702441, 'coherence_score': 0.10975845435170961,
'syn_similarity': 0.6651885329087415, 'ai_coverage': 0.1017780212466498}
WorkerID: A3DS5B06ZCD3E3, Weights: {'tone_similarity': 0.171574115892231, 'pos
_similarity': 0.007414056394940938, 'coherence_score': 0.36341440870358166, 's
yn_similarity': 0.3919072800345759, 'ai_coverage': 0.06569013897467052}
WorkerID: A1E235KE3CSO7H, Weights: {'tone_similarity': 0.022967980703954647,
'pos_similarity': 0.4233511883909377, 'coherence_score': 0.08858427221606117,
'syn_similarity': 0.20291739315959031, 'ai_coverage': 0.2621791655294561}
WorkerID: APRZ7BR8C0ZMQ, Weights: {'tone_similarity': 0.13899707332637212, 'po
s_similarity': 0.20297080252165775, 'coherence_score': 0.3339897301763226, 'sy
n_similarity': 0.21150487509151905, 'ai_coverage': 0.11253751888412848}
WorkerID: A2RUHO7I7Y4XFA, Weights: {'tone_similarity': 0.12740116336666463, 'p
os_similarity': 0.2526044692538695, 'coherence_score': 0.16992316099736468, 's
yn_similarity': 0.21759615226069645, 'ai_coverage': 0.23247505412140476}
```

```
WorkerID: A3VEF4M5FIN7KH, Weights: {'tone_similarity': 0.015233311996387014,
'pos_similarity': 0.3396012772535785, 'coherence_score': 0.4051300126429759,
'syn_similarity': 0.05936826420287267, 'ai_coverage': 0.18066713390418598}
WorkerID: A173MXK429XAZQ has only one class in target variable. Skipping...
WorkerID: A173MXK429XAZQ, Weights: {}
WorkerID: A143XRCI1YXAFE, Weights: {'tone_similarity': 0.0858856089183042, 'po
s_similarity': 0.17462276353294634, 'coherence_score': 0.15524800626333365, 's
yn_similarity': 0.205605737556591, 'ai_coverage': 0.37863788372882495}
WorkerID: A6KOTWP7N7RLU has only one class in target variable. Skipping...
WorkerID: A6KOTWP7N7RLU, Weights: {}
WorkerID: A23EWFNNOUS10B, Weights: {'tone_similarity': 0.3048916745017019, 'po
s_similarity': 0.34475784411316984, 'coherence_score': 0.020857619684989066,
'syn_similarity': 0.059117862536946245, 'ai_coverage': 0.2703749991631929}
WorkerID: A1PBRKFHSF1OF8, Weights: {'tone_similarity': 0.023120184558012544,
'pos_similarity': 0.4766552670353076, 'coherence_score': 0.221173790482766, 's
yn_similarity': 0.20222051306126332, 'ai_coverage': 0.07683024486265042}
WorkerID: A377LTGWJKY2IW, Weights: {'tone_similarity': 0.040157722294385745,
'pos_similarity': 0.1543331686820255, 'coherence_score': 0.16503382583562398,
'syn_similarity': 0.17263208026810753, 'ai_coverage': 0.4678432029198572}
WorkerID: A3HE29W5IDR394 has only one class in target variable. Skipping...
WorkerID: A3HE29W5IDR394, Weights: {}
WorkerID: AM2KK02JXXW48, Weights: {'tone_similarity': 0.032160230021922476, 'p
os_similarity': 0.2136044880029577, 'coherence_score': 0.15157906027576076, 's
yn_similarity': 0.2956836852934812, 'ai_coverage': 0.30697253640587785}
WorkerID: A17AF42SNQNH9C, Weights: {'tone_similarity': 0.24175156155020944, 'p
os_similarity': 0.30497016127870274, 'coherence_score': 0.08900057903431974,
'syn_similarity': 0.030976207180510393, 'ai_coverage': 0.33330149095625766}
WorkerID: A394JO4NEPCY3M has only one class in target variable. Skipping...
WorkerID: A394JO4NEPCY3M, Weights: {}
```

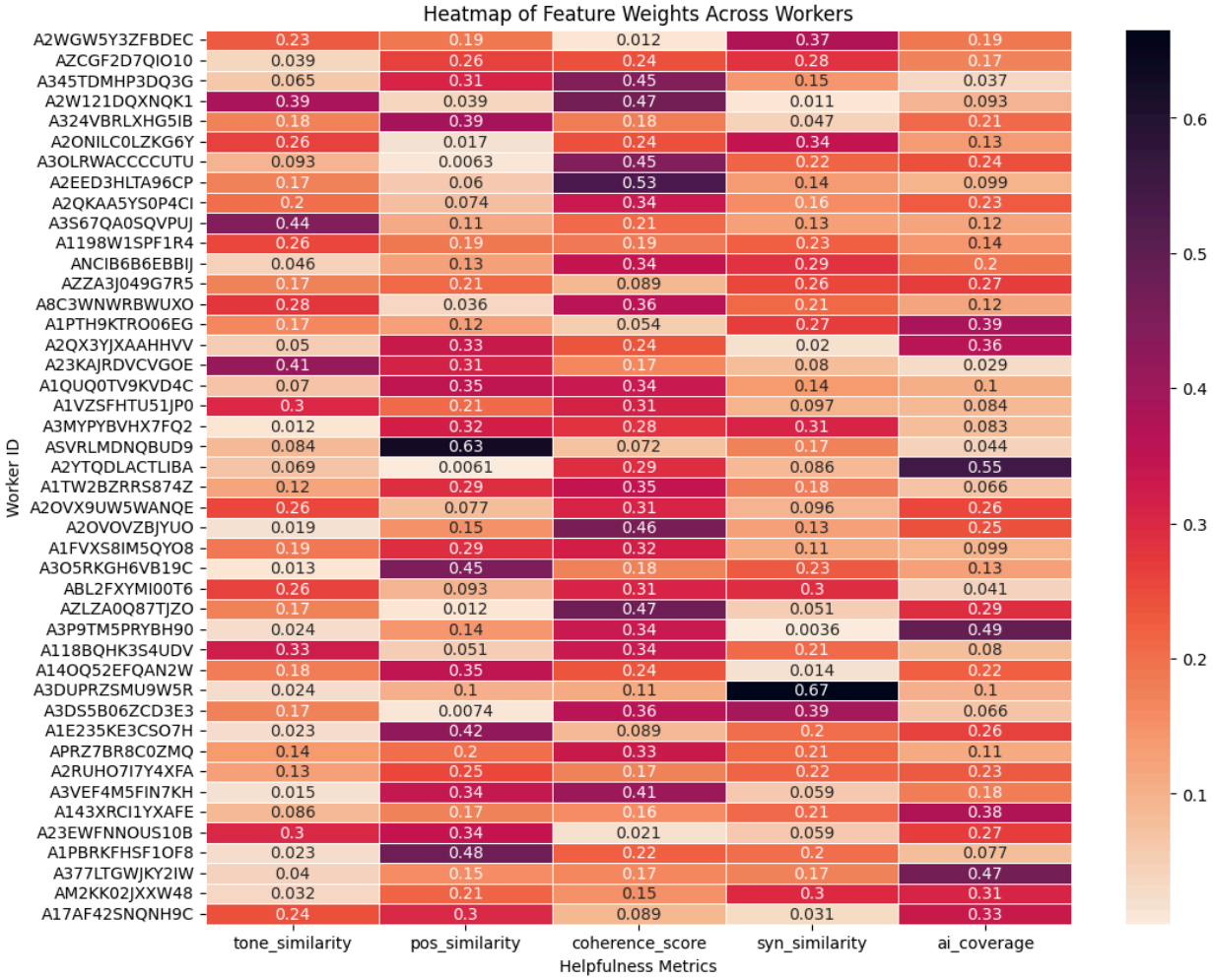
In [5]:

```python
def calculate_weights_for_all_workers(df, unique_workers):
  '''
  Calculate the weight of each helpfulness metric for all writers.

  Input: csv file

  Output: dictionary of dictionary of feature importance
  '''

  worker_weights = {}
  for worker in unique_workers:
      weights = calculate_metric_weights(df, worker_id=worker)
      worker_weights[worker] = weights

  return worker_weights
```

In [6]:

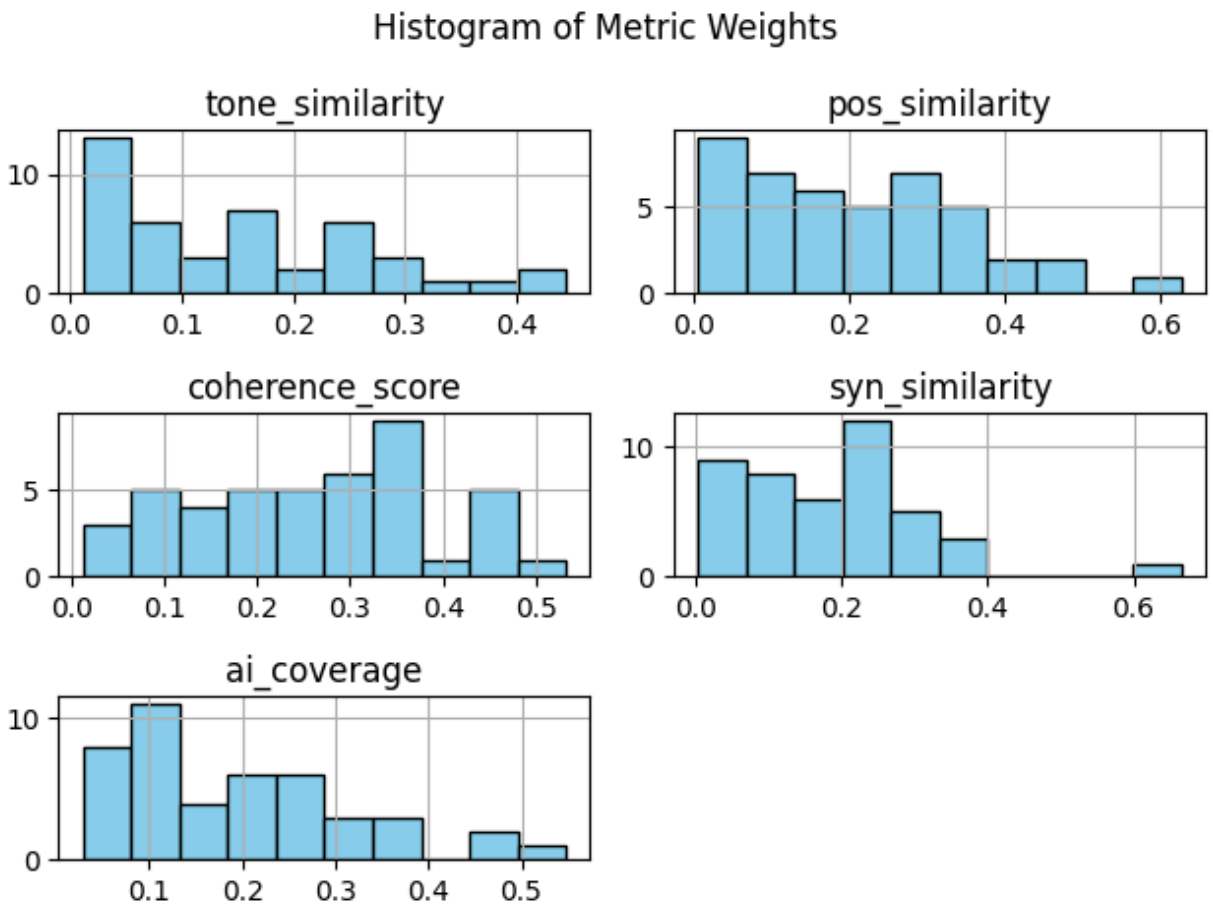
```python
import seaborn as sns
import matplotlib.pyplot as plt

worker_weights = calculate_weights_for_all_workers(df, unique_workers)
df_weights = pd.DataFrame.from_dict(worker_weights, orient='index')

plt.figure(figsize=(12, 10))
sns.heatmap(df_weights, annot=True, cmap="rocket_r", linewidths=0.5)
plt.xlabel("Helpfulness Metrics")
plt.ylabel("Worker ID")
plt.title("Heatmap of Feature Weights Across Workers")

plt.savefig('heatmap.png', dpi=300)
plt.show()
```

```
WorkerID: A17Q4QN6UE0EZC has only one class in target variable. Skipping...
WorkerID: AFIK3VBMMX6G6 has only one class in target variable. Skipping...
WorkerID: A1WKF2VH7TV0H2 has only one class in target variable. Skipping...
WorkerID: A173MXK429XAZQ has only one class in target variable. Skipping...
WorkerID: A6KOTWP7N7RLU has only one class in target variable. Skipping...
WorkerID: A3HE29W5IDR394 has only one class in target variable. Skipping...
WorkerID: A394JO4NEPCY3M has only one class in target variable. Skipping...
```

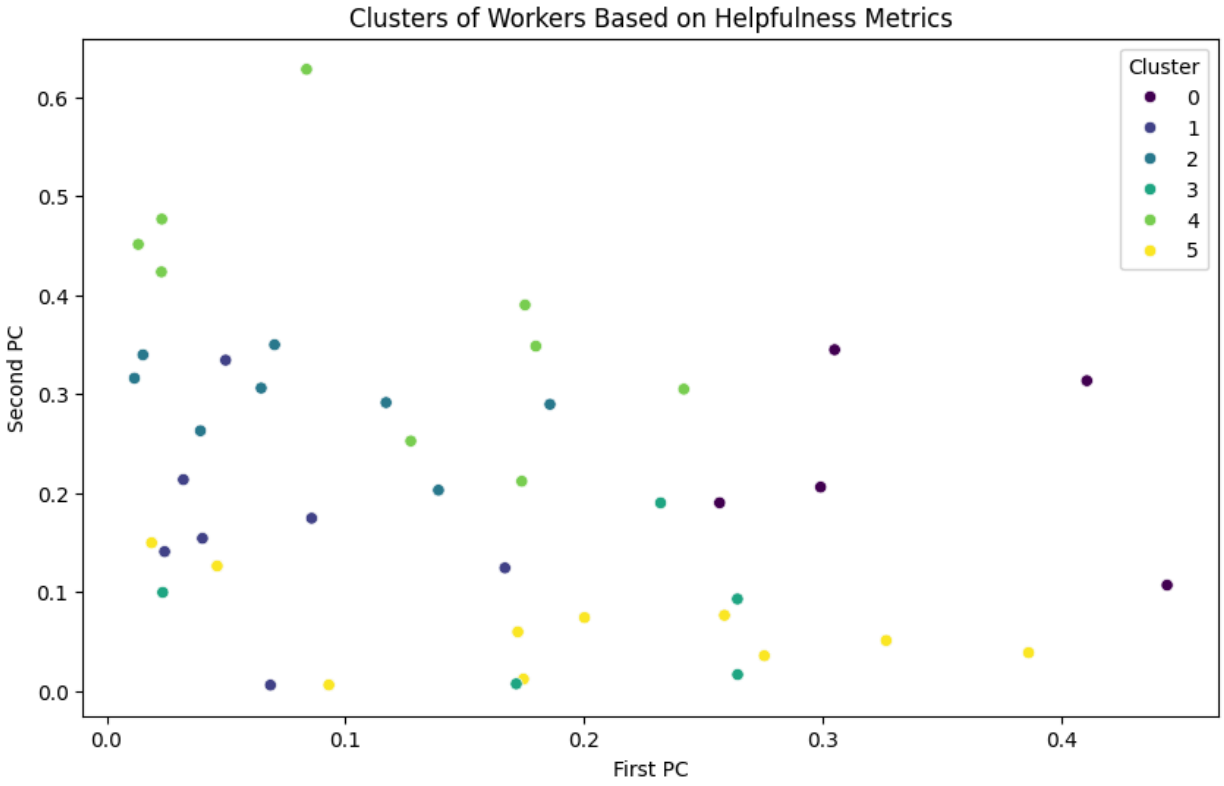


In [7]:
```
# histogram of metric weights
plt.figure(figsize=(4, 10))
df_weights.hist(bins=10, color='skyblue', edgecolor='black')
plt.suptitle('Histogram of Metric Weights', fontsize=12)
plt.xlabel('Weight', fontsize=10)
plt.ylabel('Frequency', fontsize=10)

#save img
plt.savefig('metric_weights_histogram.png')

plt.tight_layout()
plt.show()
```

```
<Figure size 400x1000 with 0 Axes>
```



In [8]:
```
from sklearn.cluster import KMeans
# Apply KMeans clustering
kmeans = KMeans(n_clusters=6, random_state=42)
df_weights['Cluster'] = kmeans.fit_predict(df_weights)

plt.figure(figsize=(10, 6))
sns.scatterplot(x=df_weights.iloc[:, 0], y=df_weights.iloc[:, 1], hue=df_weigh
plt.xlabel("First PC")
plt.ylabel("Second PC")
plt.title("Clusters of Workers Based on Helpfulness Metrics")
plt.legend(title="Cluster")
plt.show()
```


Clusters of Workers Based on Helpfulness Metrics
Cluster
0
1
2
3
4
5
Second PC
First PC
0.0
0.1
0.2
0.3
0.4
0.5
0.6